Action understanding as inverse planning

Action Understanding as Inverse Planning

Chris L. Baker, Joshua B. Tenenbaum & Rebecca R. Saxe
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Abstract

Humans are adept at inferring the mental states underlying other agents' actions, such as goals, beliefs, desires, emotions and other thoughts. We propose a computational framework based on Bayesian inverse planning for modeling human action understanding. The framework represents an intuitive theory of intentional agents' behavior based on the *principle of rationality*: the expectation that agents will plan approximately rationally to achieve their goals, given their beliefs about the world. The mental states that caused an agent's behavior are inferred by inverting this model of rational planning using Bayesian inference, integrating the likelihood of the observed actions with the prior over mental states. This approach formalizes in precise probabilistic terms the essence of previous qualitative approaches to action understanding based on an "intentional stance" (Dennett, 1987) or a "teleological stance" (Gergely et al., 1995). In three psychophysical experiments using animated stimuli of agents moving in simple mazes, we assess how well different inverse planning models based on different goal priors can predict human goal inferences. The results provide quantitative evidence for an approximately rational inference mechanism in human goal inference within our simplified stimulus paradigm, and for the flexible nature of goal representations that human observers can adopt. We discuss the implications of our experimental results for human action understanding in real-world contexts, and suggest how our framework might be extended to capture other kinds of mental state inferences, such as inferences about beliefs, or inferring whether an entity is an intentional agent.

Introduction

A woman is walking down the street, when suddenly she pauses, turns, and begins running in the opposite direction. Why? Is she just acting erratically on the way to her eventual goal? Did she change her mind about where she was going? Or did she complete an errand unknown to us (perhaps dropping off a letter in a mailbox) and rush off to her next goal? These inferences, despite their ordinariness, reveal a remarkable aspect of human cognition: our ability to infer the complex, richly-structured mental states that underlie others' actions, given only sparse observations of their behavior.

Human social interaction depends on our ability to understand and predict other people's actions in terms of the psychological states that produce behavior: chiefly, beliefs and desires. Much like visual perception, action understanding proceeds unconsciously and effortlessly but is the result of sophisticated computations that effectively solve an ill-posed, inductive problem, working backwards from sparse data to rich representations of the underlying causes. Our goal in this paper is to elucidate the computations involved in human action understanding through a combination of computational modeling and behavioral experiments. We will describe some of the first models that can explain how people perform these inferences so successfully, and that can also predict with surprising quantitative accuracy the judgments that people make.

Vision is often said to be a kind of "inverse graphics", where graphics describes the causal physical process by which images are formed from scenes. Similarly, action understanding can be characterized as a kind of "inverse planning" or "inverse reinforcement learning" (Ng & Russell, 2000). Just as computer graphics is based on mathematical models of image formation, mathematical accounts of planning and reinforcement learning have been developed by economists, computer scientists, psychologists and neuroscientists (Bellman, 1957; Watkins, 1989; Sutton & Barto, 1998; W. Schultz et al., 1997), which provide rational models of how agents should choose sequences of actions, given their goals, their prior experience, and their model of the world. Explaining an agent's actions in terms of mental states requires inverting a model of its planning process, or inverse planning: working backwards to infer the desires and beliefs that caused the agent's behavior.

Formalisms for solving the forward problems of planning and reinforcement learning are often divided into model-based and model-free approaches (Sutton & Barto, 1998; Doya, 1999; Daw et al., 2005), and there is evidence that the brain has systems corresponding to both (W. Schultz et al., 1997; Dickinson, 1985). We propose that the same kinds of cognitive machinery that support learning goal-directed action in the model-based approach – the ability to build models of the world and plan reward-maximizing sequences of actions over them – can be used in an inverse direction to infer the goals behind other agents' observed behavior.

Philosophers and psychologists have long considered non-formal versions of this proposal in discussions about "belief-desire psychology". Fig. 1(a) illustrates a typical example: a folk theory that specifies intentional agents' beliefs and desires as the causes of their behavior (*c.f.* Dennett, 1987; Wellman, 1990; Perner, 1991; Gopnik & Meltzoff, 1997). Dennett (1987) argues that this causal relation is governed by the *principle of rationality*: the expectation that intentional agents will tend to choose actions that achieve their desires most efficiently, given their beliefs about the world. At a qualitative level, inverse planning is simply running the principle of rationality in reverse. Considered as a formal computation, however, inverse planning is significantly more difficult than forward planning. Just as in vision (Barrow & Tenenbaum, 1981; Richards et al., 1996), the inverse problem is ill-posed. Its solution requires strong prior knowledge of the structure and content of agents' mental states, and the ability to search over and evaluate a potentially very large space of possible mental state interpretations. Implementing a formal version of this account, and quantitatively evaluating it with human behavioral judgments is the main contribution of our work here.

Previous experimental evidence suggests that even preverbal infants' interpretations

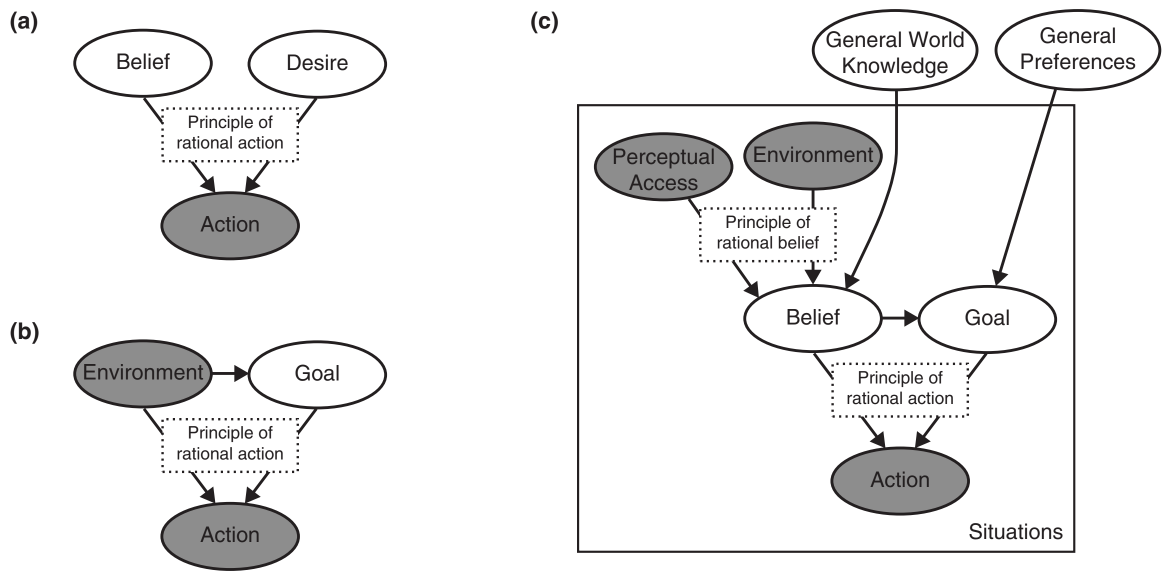

Figure 1. Modeling intuitive theories of intentional action. Diagrams use causal graph notation. Shaded nodes represent observed variables; unshaded nodes represent latent variables whose values must be inferred. Dotted boxes indicate the causal relation between variables. In this example, the observations are an agent's Action in some Environment and the agent's Perceptual Access to the value of the Environment. Given this evidence, the agent's Belief and Goal must be inferred. **(a)** Non-formal "belief-desire psychology" account of folk theories of intentional action. Non-formal accounts of human action understanding typically assume some version of this causal structure, and define the causal relation between beliefs, desires and actions in terms of qualitative, context-specific commonsense rules (*e.g.* Wellman & Bartsch, 1988; Fodor, 1992). **(b)** A simplified version of (a) proposed by Gergely et al. (1995) and Csibra & Gergely (1997) as a model of infants' early-developing action understanding competency. We formalize this intuitive theory as Bayesian inverse planning. The functional form of the causal relation between Environment, Goal and Action is given by rational probabilistic planning in Markov decision problems, and goal inference is performed by Bayesian inversion of this model of planning. **(c)** A generalized framework for human action understanding. This relates the non-formal account in (a) to the formal model in (b) by sketching how agents' Beliefs and Desires (or Goals) depend on their Perceptual Access to the Environment, mediated by their General World Knowledge and General Preferences. The model in (b) is a limiting case of (c), in which agents are assumed to have complete Perceptual Access to the Environment, constraining their Beliefs to be equal to the Environment.

of behavior are qualitatively consistent with the inverse planning view (Meltzoff, 1995; Gergely et al., 1995; Meltzoff, 1988; Gergely et al., 2002; Csibra et al., 2003; Sodian et al., 2004; Phillips & Wellman, 2005). Six-month-old infants interpret simple human motions as goal-directed actions, and expect that subsequent behavior will be consistent with these inferred goals (Woodward, 1998). That is, when actions *could* be interpreted as a rational or efficient means to achieve a concrete goal, infants expect the actor to continue to use the most efficient means to achieve the same goal, even when the environment changes. Gergely, Csibra and colleagues found that six- to twelve-month old infants extend the same expectations to the novel (and relatively impoverished) movements of two-dimensional shapes (Gergely et al., 1995; Csibra et al., 1999, 2003). In this context, infants' inferences were flexible and productive: given information about any two of the environment, the action and the goal, infants could infer the likely value of the third. To account for these findings, Gergely et al. (1995) proposed an early-developing, non-mentalistic version of Fig. 1(a), shown in Fig. 1(b). On their account, the Environment represents concrete situational constraints on the agent's available actions, such as the agent's own location and the location of other agents, objects, or obstacles, and the Goal is some point or entity in the Environment. Together, the Environment and Goal provide a basis for the agent's Action under a simple version of the rationality principle known as the *teleological stance*. Gergely et al. (1995) argue that this simplified schema forms the core of a more sophisticated, later-developing mentalistic theory of intentional action.

This research, along with the essential computational difficulty of action understanding, raises several open questions about how action understanding works in the mind. Can human action understanding competency be described by formal models, or is our intuitive psychological knowledge vague and heterogeneous? If action understanding can be formalized, can people's judgments be explained by models of inverse planning? Does inverse planning explain people's judgments better than simple heuristic alternatives? If human judgments are best explained by inverse planning, what is the form and content of our representations of agents' mental states and actions – the priors that make inductive mental state inferences possible?

To address these questions, we formalize action understanding as a Bayesian inference problem. We model the intuitive causal relation between beliefs, goals and actions as rational probabilistic planning in Markov decision problems (MDPs), and invert this relation using Bayes' rule to infer agents' beliefs and goals from their actions. We test our framework with psychophysical experiments in a simple setting that allows us to collect a large amount of fine-grained human judgments to compare with the strong quantitative predictions of our models.

Specifically, we use the tools of Bayesian inverse planning to formalize the action understanding schema shown in Fig. 1(b). Inspired by Gergely et al. (1995), we assume that the agent's Action depends directly on the Environment and the Goal, without requiring a separate representation of the agent's beliefs. To specify the agent's likely Actions as a function of the constraints of the Environment and the agent's Goal, these variables are encoded within an MDP, and the causal relation between them is computed by a mechanism for rational planning in MDPs. We assume that the planning relation is probabilistic, tolerating a certain amount of noise or variability in how agents can execute their plans.

Fig. 1(c) sketches a more general intuitive theory of rational action, intended to capture various qualitative proposals in the theory of mind literature (*e.g.* Wellman & Bartsch (1988); Wellman (1990); Bartsch & Wellman (1995); see also Goodman et al. (2006) for a related formal account). This schema extends Fig. 1(a) by describing how beliefs depend on perceptual access to the environment, mediated by general world knowledge, and how goals depend on general preferences over states of the world. General world knowledge and preferences are high-level variables that apply across situations, while new beliefs and goals are generated specifically for each situation. The specific models we work with in this paper (Fig. 1(b)) correspond to the special case in which agents are assumed to have full perceptual access to the environment, thereby constraining the contents of their beliefs to be equal to the environment. A formal implementation of the more general framework in

Fig. 1(c) is beyond our scope here, but in the General Discussion we consider the additional computational assumptions needed to extend our work in that direction, to allow reasoning about the unknown contents and origins of agents' beliefs.

The Bayesian inversion of MDP models of behavior requires strong priors over the space of agents' goals. In our framework, the most basic concept of a goal corresponds to the objective to bring about a particular state of the environment. However, this is clearly too inflexible to describe the sophisticated kinds of goals that humans can attribute to other agents, and there are many ways that the basic goal concept can be extended. As a first step, in this paper we consider two extensions to the most basic goal concept, which roughly correspond to the explanations of the woman's behavior in the introductory vignette: goals that can change over time and goals with more complex content, such as subgoals along the way to a final goal. We also formulate a simple heuristic alternative based on low-level motion cues as a limiting case of the changing-goal prior. We describe Bayesian inverse planning models based on these different goal priors in the Computational Framework section, and compare how accurately they predict people's judgments in our experiments.

Our experiments use a stimulus paradigm of animated displays of agents moving in simple maze-like environments to reach goal objects, inspired by stimuli from many previous studies with children and adults (*e.g.* Heider & Simmel, 1944; Gergely et al., 1995; R. Gelman et al., 1995; Scholl & Tremoulet, 2000; Tremoulet & Feldman, 2000; Zacks, 2004; R. T. Schultz et al., 2003; J. Schultz et al., 2005; Tremoulet & Feldman, 2006). This paradigm allows fine-grained experimental control of agents' actions, environment and plausible goals, and is ideal for both psychophysical experiments and computational modeling. Although this methodology greatly simplifies real-world action understanding, these kinds of stimuli evoke a strong sense of agency and the impression of mental states to adults (Heider & Simmel, 1944; Tremoulet & Feldman, 2000, 2006) (even when adult subjects are instructed *not* to make mentalistic interpretations (Heberlein & Adolphs, 2004)), and can lead to the formation of expectations consistent with goal-directed reasoning in infants (Gergely et al., 1995; Csibra et al., 1999, 2003). There is evidence that these kinds of stimuli recruit brain regions associated with action perception in adults (Castelli et al., 2000; R. T. Schultz et al., 2003; J. Schultz et al., 2005), suggesting a common mechanism with real-world action understanding. Further, these stimuli can represent quite complex situations and events (Heider & Simmel, 1944), with similar abstract structure to more naturalistic contexts. Similarly, our computational models can be extended to much more general contexts than the simple scenarios in our experiments, as we will show with several examples in the Computational Framework section.

We present three experiments, which measure people's online goal inferences, retrospective goal inferences, and prediction of future actions based on previous goal inferences, respectively. Taken together, our experiments test whether human action understanding in our experimental domain can be explained by inverse planning. Individually, our experiments probe the space of representations that people apply in action understanding. Each experiment includes special conditions to distinguish the predictions of inverse planning models based on different goal priors. By comparing which of these models produces inferences that match people's judgments most accurately in each experimental context, we show how our approach can be used to elucidate the prior knowledge applied in human action understanding.

Computational Framework

Our computational framework formalizes action understanding as Bayesian inverse planning: the Bayesian inversion of models of probabilistic planning in Markov decision problems (MDPs). This section will provide an overview of our framework and its application to our experimental stimuli. First, we will describe the encoding of the maze-world scenarios of our experiments into MDPs. We will also sketch the MDP encoding of several more realistic environments and contexts than those of our experiments to emphasize the generality of Bayesian inverse planning principles. Next, we will describe the computations underlying the mechanism for planning in MDPs. We will then sketch the Bayesian computations involved in inverting MDP models of planning, and give examples of the kinds of structured goal priors that are required to perform these computations. Finally, we will compare our framework with previous models of action understanding. Our overview in this section will be fairly high-level, with the formal details provided in a separate appendix.

Our framework uses MDPs to capture observers' mental models of intentional agents' goal- and environment-based planning. MDPs are a normative framework for modeling sequential decision making under uncertainty, widely used in rational models of human planning and reinforcement learning (Dayan & Daw, 2008), and in real-world applications in operations research and other fields (Feinberg & Shwartz, 2002; Puterman, 2005). An MDP represents an agent's model of its interaction with its environment. MDPs encode all relevant information about the configuration of the world and the agent with the *state* variable. MDPs also represent the affordances of the environment: what actions the agent can take and a causal model of how these actions change the state of the world. Finally, MDPs represent the subjective rewards or costs caused by the agent's actions in each state.

In our maze-world scenarios, the state includes the location of all obstacles, potential goals and other objects, and the location of the agent. Agents can take 9 different actions: Stay, North, South, East, West, NorthEast, NorthWest, SouthEast and SouthWest, except when these actions are blocked by obstacles. For simplicity, we assume that actions always lead to their intended movements. The agent's goal is to achieve a particular state of the world, and each action is assumed to produce a small cost to the agent until this goal is reached. Once the agent reaches its goal it is satisfied, and these costs cease. We define costs to be proportional to the negative distance of the intended movement: actions North, South, East, and West have costs proportional to -1, and actions NorthEast, NorthWest, SouthEast and SouthWest have costs proportional to $-\sqrt{2}$. We also define the cost of the Stay action to be proportional to -1 to capture the desire for continual progress toward the goal. Formally, these assumptions induce a class of stochastic shortest path problems (Bertsekas, 2001), implying that rational agents should plan to reach their goal as quickly and efficiently as possible.

To illustrate the application of MDPs to another domain, consider the game of golf. In golf, the state is comprised by the current hole, the current score, the current stroke number, and the position of the ball. Actions must specify club selection and the type of shot to be attempted. The causal model of the effect of actions reflects the inherent uncertainty about where a shot ends up, with the outcome of difficult shots being more uncertain than others. In golf, each shot has a cost of 1, and rational players try to minimize their score.

In addition to golf, many other games can be modeled as MDPs, such as blackjack (Sutton & Barto, 1998), backgammon (Tesauro, 1994), Tetris (Bertsekas & Tsitsiklis, 1996), and football (Bertsekas & Tsitsiklis, 1996).

As another example, consider the job of the head chef of a restaurant. Given a menu, the chef's goal is to prepare each dish as well and as quickly as possible to maximize customer satisfaction and restaurant capacity. The state consists of the number of kitchen staff and the available ingredients, burners, ovens, cooking implements, counter space, *et cetera*. Actions include delegation of tasks to kitchen staff, the allocation of cooking resources to different dishes, and the chef's own hands-on preparation of dishes. As in golf, the causal model reflects the uncertainty in preparation time and the performance of staff members. Relative to golf, cooking has rich logical and hierarchical structure, where every step has a large number of pre- and post-conditions, and individual actions may enable multiple subsequent steps. For instance, sending a line cook after two raw eggs furthers both the goal to cook a frittata as well the goal to prepare a soufflé.

Planning formally describes the way that intentional agents choose actions to achieve their goals in an MDP. An optimal plan is one that provides the minimum-expected-cost course of action to the goal from every state. Our models assume that agents choose the optimal action only probabilistically. This yields a probability distribution over $\mathsf{Actions}$, given a Goal and the $\mathsf{Environment}$, denoted $P(\mathsf{Actions}|\mathsf{Goal},\mathsf{Environment})$, which provides the functional form of the probabilistic planning relation in Fig. 1(b). We assume that an agent's actions are distributed in proportion to the softmax function of the expected value (negative expected cost) of each available action. The level of determinism in the agent's actions is represented by the parameter β: higher β values yield greater determinism (less noise), and lower β values yield less determinism (more noise). We describe an algorithm for probabilistic planning based on dynamic programming (Bellman, 1957) in the appendix.

Given an MDP model of goal-directed planning, Bayesian inverse planning computes the posterior probability of a Goal, conditioned on observed $\mathsf{Actions}$ and the $\mathsf{Environment}$, using Bayes' rule:

$$P(\mathsf{Goal}|\mathsf{Actions},\mathsf{Environment}) \;\propto\; P(\mathsf{Actions}|\mathsf{Goal},\mathsf{Environment})P(\mathsf{Goal}|\mathsf{Environment}). \quad (1)$$

In this equation, $P(\mathsf{Actions}|\mathsf{Goal},\mathsf{Environment})$ is the likelihood of the Goal given observed $\mathsf{Actions}$ and the $\mathsf{Environment}$, defined above as probabilistic planning in an MDP. $P(\mathsf{Goal}|\mathsf{Environment})$ is the prior probability of the Goal given the $\mathsf{Environment}$, which sets up a hypothesis space of goals that are realizable in the environment. Inverse planning integrates bottom-up information from observed actions and top-down constraints from the prior to infer the Goal, given observed $\mathsf{Actions}$ and the $\mathsf{Environment}$. We describe inverse planning models based on several different goal priors below.

Inverse planning also enables goal-based prediction of future actions in novel situations, given prior observations of behavior in similar situations. For simplicity, in this paper, our goal-based prediction experiment presents scenarios where agents' environments and goals remain constant across observations (although our framework handles cases where the environment changes as well (Baker et al., 2006)). Action prediction averages over the probability of possible future $\mathsf{Actions}'$, given a Goal and the $\mathsf{Environment}$, weighted by the posterior over the Goal given previously observed $\mathsf{Actions}$ and the $\mathsf{Environment}$ from Equation 1. This is just the posterior predictive distribution over future $\mathsf{Actions}'$ given previous

Actions and the Environment:

$$P(\mathsf{Actions}'|\mathsf{Actions}, \mathsf{Environment}) = \sum_{\mathsf{Goal}} P(\mathsf{Actions}'|\mathsf{Goal}, \mathsf{Environment}) P(\mathsf{Goal}|\mathsf{Actions}, \mathsf{Environment}). \quad (2)$$

We use Equations 1 and 2 to model people's judgments in our experimental tasks. In Experiment 1, we model people's online goal inferences using an online version of Equation 1. In Experiment 2, we model people's retrospective goal inferences using a smoothed version of Equation 1. In Experiment 3, we model people's predictions of agents' future actions, given observations of their past behavior of varying complexity, using Equation 2.

We formulate several inverse planning models, based on different forms of the prior over goals, which we refer to as M1, M2 and M3. Each model specifies a family of goal priors depending on one or a few continuous parameters. These models are surely much too simple to capture the full range of goals that people can attribute to an intentional agent, and they simplify in different ways. Each model should be thought of as just a first approximation to some aspects of people's goal priors, embodying certain abstract principles that could be important (along with many others) in structuring our expectations about intentional action. By comparing the predictions of these models to people's judgments from our experiments, we test the extent to which they capture significant dimensions of people's prior knowledge, and in what contexts. We sketch the goal priors used by M1, M2 and M3 below. The formal details of these goal priors are provided in the appendix, including derivations of Equations 1 and 2 specific to each model.

The first inverse planning model we consider (M1) assumes that a goal refers to a single state of the environment that an agent pursues until it is achieved. Given a goal, probabilistic planning produces actions that tend to move the agent closer to the goal state, but depending on β, the level of determinism, planning can sometimes yield unexpected actions such as changes in direction or detours. Higher β values will fit actions that follow the shortest path very well, but will fit noisy action sequences very poorly. Lower β values will fit most action sequences moderately well, but will not fit any action sequence particularly closely. For example, in the introductory vignette the best explanation of the woman's erratic behavior in terms of M1 is that she tends to pursue her goals in a particularly noisy manner, or has a low β value. Moreover, the assumption of a noisy (low β) agent is the *only* way that M1 can explain paths that deviate from the simplest notion of rational action, described by a shortest path to a single, fixed goal. Our other models also support such explanations; formally, they have the same β parameter representing the agent's level of determinism. They differ in allowing a broader range of alternative explanations based on richer representations of agents' possible goals.

The second inverse planning model we consider (M2) is an extension of M1 that assumes that agents' goals can change over the course of an action sequence. This allows M2 to explain changes in direction or indirect paths to eventual goals, as in the attribution that the woman in the introductory vignette had changed her mind about where she was headed. M2 represents the prior probability that an agent will change its goal after an action with the parameter γ. With $\gamma = 0$, goal changing is prohibited, and M2 is equivalent to M1. With γ close to 0, the model rarely infers goal changes, implying that all past and recent actions are weighted nearly equally in the model's goal inferences. With γ close to 1, the model infers goal changes frequently, and only the most recent actions factor into the model's goal inferences. Intermediate values of γ between 0 and 1 interpolate between these extremes, adaptively integrating or forgetting past information depending on the degree of evidence for subsequent goal changes.

The third inverse planning model we consider (M3) is an extension of M1 that assumes that agents can have subgoals along the way to their final goal. For example, M3 can capture the inference that the woman in the introductory vignette wanted to first complete a task (such as dropping off a letter) before pursuing her next goal. M3 represents the prior probability of a subgoal with the parameter κ. With $\kappa = 0$, subgoals are prohibited, and M3 is equivalent to M1. With $\kappa > 0$, M3 can infer a sequence of subgoals to explain more complex behaviors, such as paths with large detours from the shortest path to a final goal.

Finally, we consider a simple alternative heuristic (H) based on low-level motion cues, inspired by Blythe et al. (1999), Zacks (2004) and Barrett et al. (2005). H looks only at an agent's most recent action, rather than a whole sequence of actions, assuming that at any given time, the agent's goal is probably the object toward which it had most recently moved. Including H in our comparisons allows us to test whether the full machinery of inverse planning is needed to explain human goal inferences, and in particular, the extent to which temporal integration over agents' entire paths is an essential feature of these inferences. For the sake of comparison with our inverse planning models, we formulate H as a special case of M2 in which agents' goals can change arbitrarily after every action, *i.e.* γ is set to its extreme value of 1. For many action sequences, both in our maze-world settings and in everyday situations, H makes similar predictions to the inverse planning models. These are cases that support a single unambiguous goal interpretation through the action sequence. However, our experiments are designed to include a subset of conditions with more complex trajectories that can distinguish between H and inverse planning models.

Related Work

Previous computational models of action understanding differ from our framework along several dimensions. Much of the classic work on action understanding relies on logical representations of the domain and the agent's planning process (Schank & Abelson, 1977; Kautz & Allen, 1986). These approaches use sophisticated, hierarchical representations of goals and subtasks, such as scripts and event hierarchies, to model the structure of agents' behavior, and model goal inference in terms of logical sufficiency or necessity of the observed behavior for achieving a particular goal. Probabilistic versions of these ideas have also been proposed, which allow inductive, graded inferences of structured goals and plans from observations of behavior (Charniak & Goldman, 1991; Bui et al., 2002; Liao et al., 2004). However, these approaches assume that the distribution over actions, conditioned on goals, is either available *a priori* (Charniak & Goldman, 1991; Bui et al., 2002), or must be estimated from a large dataset of observed actions (Liao et al., 2004). An alternative is to model the abstract principles underlying intentional action, which can be used to generate action predictions in novel situations, without requiring a large dataset of prior observations. Various forms of the rationality assumption have been used to achieve this in both logical and probabilistic models of action understanding (Kautz & Allen, 1986; Ng & Russell, 2000; Verma & Rao, 2006). However, these models have not compared against human judgments, and have not explored the kinds of structured goal representations necessary to explain human action understanding. In this paper, we integrate probabilistic models of rational planning with simple structured representations of agents' goals to model human action understanding. Although we do not directly test any of the models described above, we test whether their computational principles can account for human goal inferences in our experiments.

Experiment 1

Our first experiment measured people's online goal inferences in response to animated stimuli of agents moving to reach goal objects in simple maze-like environments. Our stimuli varied the environmental context, including the configuration of marked goals and obstacles, and varied agents' paths and the point at which participants' judgments were collected. This yielded fine-grained temporal measurements of human goal inferences and their sensitivity to various actions and contexts. We addressed the motivating questions from the Introduction by comparing how accurately M1, M2, M3 and H predicted participants' judgments. Comparing models based on different goal priors revealed aspects of the form and content of the prior knowledge underlying human goal inferences. Comparing the accuracy of these models with H tested whether people's judgments in our experimental domain were best explained as a process of inverse planning or the application of a simple heuristic.

Method

Participants.

Participants were 16 members of the MIT subject pool, 9 female, 7 male.

Stimuli.

Subjects viewed short animations of agents moving in simple mazes. Agents were represented by small moving circles, and as they moved through the environment, traces of their trajectories trailed behind them to record their entire movement history as a memory aid. Each displayed movement sequence paused at a *judgment point*: a point in the middle of the agent's trajectory before a particular goal was achieved, where subjects reported their online goal inferences. The environment was a discrete grid of squares that agents could occupy, with dimensions of 17 squares wide by 9 squares high. Agents' movements were restricted to adjacent squares, with directions {N,S,E,W,NE,NW,SE,SW}. Known goals were displayed as capital letters "A", "B" and "C", and walls were displayed as solid black barriers. Animations were shown from an overhead perspective (*i.e.* looking down on a room with a wall in the middle). Example stimuli from Experiment 1 are shown in Fig. 2(a).

Design.

All 36 conditions of Experiment 1 are shown in Fig. 3. Our experimental design varied three factors: goal configuration, obstacle shape and agent path. There were four different goal configurations, displayed in columns 1-4 of Fig. 3. Only the location of goal C changed across conditions; goals A and B were always in the upper and lower right corners, respectively. There were two different obstacle shapes: "Solid" and "Gap". Every environment shown had a wall-like obstacle extending up from the bottom edge. In the

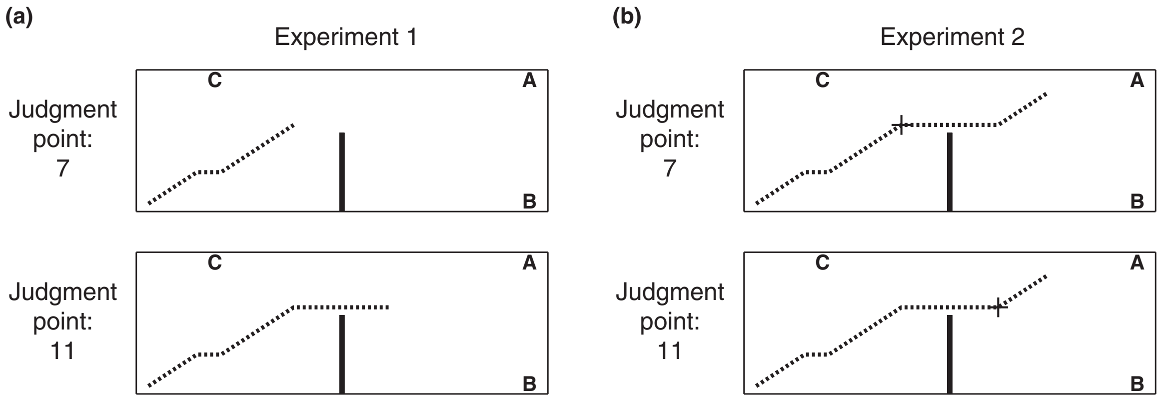

Figure 2. Stimulus paradigm for Experiments 1 and 2. Each stimulus presented an animation of an agent's path (marked by a dashed line) ending at a judgment point: a pause in the animation that allowed participants to report their online inferences of the agent's goal at that point. **(a)** Experiment 1: online goal inference task. Subjects rated how likely each marked goal was at the judgment point. **(b)** Experiment 2: retrospective goal inference task. Subjects rated how likely each marked goal was at an earlier judgment point, given by the "+" along the dashed line.

Solid conditions this wall was unbroken, while in the Gap conditions it had a hole in the middle through which the agent could pass. The first, fourth, and seventh rows of Fig. 3 represent the Solid conditions, while the remaining rows represent the Gap conditions.

Based on the goal configuration and obstacle shape, the agent's path was generated by making two choices: first, which goal (A, B or C) the agent was heading toward, and second, whether the agent went around the obstacle or through it. The second choice only applied in the Gap conditions; in the Solid conditions the agent could only move around the obstacle. In Fig. 3, paths are grouped as "A" paths, "B" paths and "C" paths, respectively. Because of C's varying location, there were 8 unique C paths, while there were just two unique A paths and two unique B paths because the locations of A and B were fixed. All paths started from the same point, marked with an "**x**" in Fig. 3.

Each condition included a number of trials, which varied the length of the path shown before a judgment was required. Different conditions queried subjects at different judgment points, selected at informative points along the paths. Fig. 2(a) displays two stimuli with judgment points of 7 and 11, respectively, as they were plotted for our subjects. In Fig. 3, many of the initial trials are identical, and only differ in their eventual destination (*e.g.* corresponding trials in rows 1 and 4 of Fig. 3 are identical up to judgment point 10). Subjects were only shown unique stimuli, and after all redundant conditions were removed, there were 99 stimuli in total, all represented in Fig. 3.

Procedure.

Participants were given a cover story to establish assumptions about our experimental scenarios, including the assumption of intentional agency, a model of agents' environments, and a hypothesis space of agents' goals. Participants were told they would be viewing videos of members of an intelligent alien species collected by scientists, and that each video displayed a different alien moving toward a different goal in the environment. They were

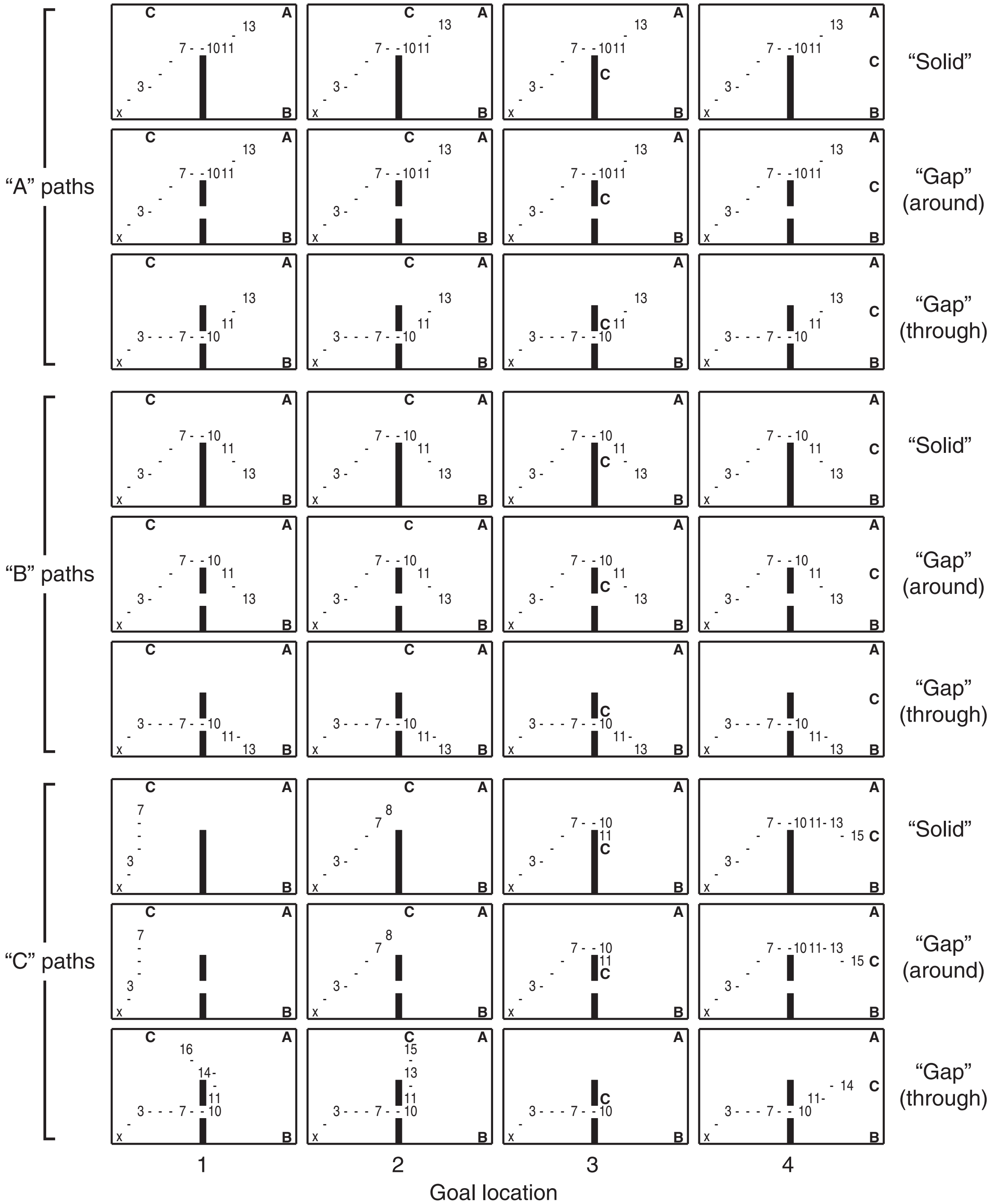

Figure 3. All stimuli from Experiment 1. We varied three factors: goal configuration, obstacle shape and agent path. There were four goal configurations, displayed in columns 1-4. Path conditions are grouped as "A" paths, "B" paths and "C" paths. There were two obstacle shape conditions: "Solid" and "Gap". There were 36 conditions, and 99 unique stimuli in total.

instructed that aliens could not pass through walls, but that they could pass through gaps in walls. They were told that after each video, they would rate which goal the alien was pursuing.

Stimulus trials were ordered with the earliest judgment points presented first to prevent hysteresis effects from showing longer trials before their shorter segments. Trials with the same judgment points were shown in random order. On each trial, the animation paused at a judgment point, allowing participants to report their online inferences of the agent's goal at that point. Subjects first chose which goal they thought was most likely (or if two or more were equally likely, one of the most likely). After this choice, subjects were asked to rate the likelihood of the other goals relative to the most likely goal, on a 9-point scale from "Equally likely", to "Half as likely", to "Extremely unlikely".

Modeling.

Model predictions take the form of probability distributions over agents' goals, given by Equation 1 (specific versions for M1, M2, M3 and H are provided in the appendix). Our models assumed that all goals were visible, given by the three marked locations in our stimuli. M3 assumed there were either 0 or 1 subgoals, which could correspond to any location in the environment. To put people's goal inferences on the same scale as model predictions, subjects' ratings were normalized to sum to 1 for each stimulus, then averaged across all subjects and renormalized to sum to 1. The within-subjects normalization guaranteed that all subjects' ratings were given equal weighting in the normalized between-subjects average.

Results

We present several analyses of how accurately M1, M2, M3 and H predicted people's online goal inferences from Experiment 1. We begin with a qualitative analysis, which compares subjects' data and model predictions from several conditions of Experiment 1 to illustrate the kinds of behavior that people are sensitive to in action understanding, and to show how closely inverse planning models captured people's judgments. We then turn to several quantitative analyses which rigorously support our previous qualitative observations and address the motivating questions from the Introduction.

Fig. 4 shows examples of our qualitative comparisons between participants' goal inferences and the predictions of M2. As we will argue quantitatively below, M2 was the model that best explained people's judgments in Experiment 1. The conditions in Fig. 4(a) were selected to highlight the temporal dynamics of subjects' ratings and model predictions in response to different observed actions and environments. Each condition in Fig. 4(a) differs from adjacent conditions by one stimulus feature, which illustrates the main effects of changing the goal configuration, the obstacle shape and the agent's path.

Participants' average ratings with standard error bars from these conditions are shown in Fig. 4(b). These examples illustrate several general patterns of reasoning predicted by our models. At the beginning of each trajectory, people tended to be uncertain about the agent's goal. As more of the trajectory was observed, their judgments grew more confident (*e.g.* compare conditions 1 and 2 of Fig. 4(b)). A comparison of adjacent conditions in Fig. 4(b) shows how changing stimulus features had specific effects on people's goal inferences – some subtle, others more dramatic.

Predictions of M2 using the best-fitting parameters are shown in Fig. 4(c). The model

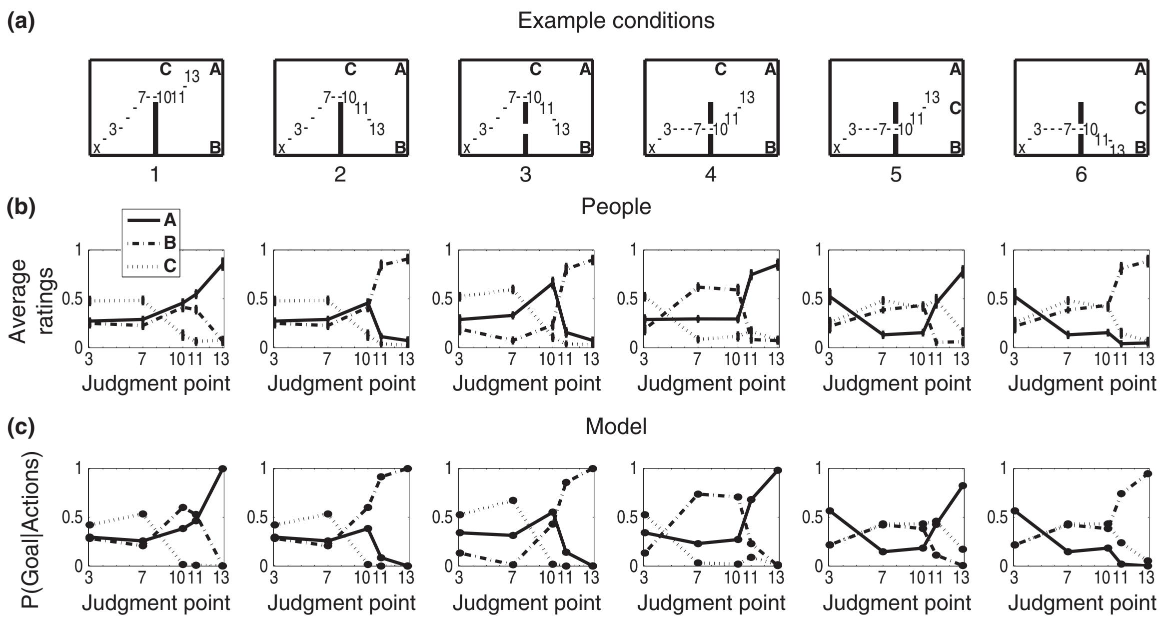

Figure 4. Example conditions, data and model predictions from Experiment 1. **(a)** Stimuli illustrating a range of conditions from the experiment. Each condition differs from adjacent conditions by one feature—either the agent's path, the obstacle, or the goal locations. **(b)** Average subject ratings with standard error bars for the above stimuli. **(c)** Predictions of inverse planning model M2 with parameters $\beta = 2.0$, $\gamma = 0.25$.

predicted people's judgments with high accuracy across all conditions, including the effects of varying key stimulus features of the agent's environment or path. M2 can be used to explain why participants' judgments varied as they did across these conditions, based on the general expectations about rational action that M2 embodies: agents will tend to move along the most efficient paths toward their goal, and goals tend to stay constant over time but occasionally can change.

Conditions 1 and 2 in Fig. 4 differed only at the end of the path, after the agent had passed the obstacle. Before this point people assigned similar probabilities to goals A and B, because the trajectory observed was essentially the shortest path toward both goals. Beyond the obstacle the agent took a step in the direction of goal A (Condition 1) or goal B (Condition 2) and people's uncertainty resolved, because the trajectory was now consistent with the shortest path to only one goal and departed significantly from the shortest path to the other.

Condition 3 in Fig. 4 differed from Condition 2 only in the presence of a gap in the obstacle. Relative to Condition 2, this resulted in a significantly lower probability assigned to goal B early on, as the agent went around the obstacle (see step 7 in particular). This can be thought of as a counterfactual inference: if the agent had been heading toward B originally, it would have gone straight through the gap rather than around the obstacle; this alternative shortcut path was not available in Condition 2. Once the agent had passed the obstacle and turned toward B, people were very quick to change their judgment to B. The model explained this as the inference that the agent's goal, while initially probably A or C, had now switched to B.

Condition 4 in Fig. 4 reversed the pattern of Condition 3: the agent initially went through (rather than around) the obstacle and then turned toward A (rather than B). Before the obstacle, people rated goal B most likely and A about half as likely, because the agent's trajectory was along the shortest path to B and followed a less efficient path to A. Once the agent passed the obstacle and turned toward A, people changed their judgment to B, which the model explained as a change in the agent's goal from B to A (or, less likely, the choice of an inefficient path to reach A).

Conditions 5 and 6 in Fig. 4 changed the position of C. Before the agent passed through the obstacle, the agent's trajectory was now along the shortest path to B and C, which increased the relative probability assigned to C and decreased the probability of A and B. After the obstacle, the agent either turned toward A (Condition 5) or B (Condition 6). In Condition 5, this conflicted with previous evidence, and people quickly changed their goal inference to A, which the model explained as a change in goals (or the choice of a less efficient path). In Condition 6, the movement after the obstacle was consistent with B, the previously inferred goal, and the probability assigned to B continued to increase.

The basic logic of our quantitative analyses was to compare how accurately different models predicted people's judgments using measures of correlation. Our overall quantitative analysis computed the correlation of each model class with people's data and assessed the statistical significance of the differences between these correlations using bootstrap cross-validation (Cohen, 1995). Bootstrap cross-validation (BSCV) is a technique for model selection, which measures the goodness-of-fit of models to data while preventing overfitting and controlling for model complexity. We describe the details of our BSCV analysis further in the appendix. The average correlations for each model from our analysis are shown in Table 1. M2 performed best, correlating significantly higher with people's judgments than M1 ($p_{BSCV} < 0.0001$), M3 ($p_{BSCV} < 0.0001$) and H ($p_{BSCV} = 0.032$). H performed second best, correlating significantly higher with people's judgments than M1 ($p_{BSCV} < 0.0001$) and M3 ($p_{BSCV} < 0.0001$). M3 correlated significantly higher with people's judgments than M1 ($p_{BSCV} = 0.0004$), and M1 performed worst.

	M1	M2	M3	H
$\langle r \rangle$	0.82 (0.017)	0.97 (0.0046)	0.93 (0.012)	0.96 (0.0027)

Table 1: Bootstrap cross-validated correlation of inverse planning models and a simple heuristic alternative with people's data from Experiment 1. Numbers in parentheses indicate standard deviation.

To illustrate the pattern of errors for each model, Fig. 5 shows scatter plots of the correspondence between participants' judgments and M1, M2, M3 and H using their best-fitting parameter values. In Fig. 5, M2 has the fewest outliers. M1 has the most outliers, and M3 and H have a significant number of outliers as well.

These results raised several questions. First, how sensitive was the degree of fit of each model to the values of the parameters? We compared participants' goal inferences to model predictions under a range of different parameter settings. Plots of the correlation of M1, M2, M3 and H with participants' data across all tested values of β, γ and κ are shown in the appendix. These plots show that a robust range of parameters around the

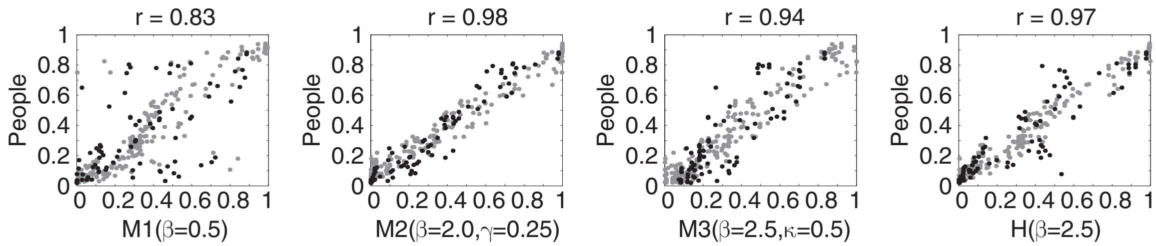

Figure 5. Scatter plots of model predictions using best-fitting parameter settings (X-axes) versus people's online goal inferences (Y-axes) for all Experiment 1 stimuli. Points from the Gap obstacle condition at judgment points 10 and 11 are plotted in black. These trials account for most of the outliers of M1, M3 and H.

best-fitting values yielded correlations that were close to the optimum for each model.

Second, what do the best-fitting parameter values tell us about how each model, with its specific goal representations, explains the range of observed actions? M1 correlated best with people's data at low values of β. In trials where the agent did not follow the shortest path to a single goal, but instead made detours around or through the obstacle, M1 could only explain these actions as noise, and low β values allowed M1 to filter out this noise by integrating information more slowly over time. Our other models performed best with higher β values than M1, because representing more sophisticated goals allowed them to assume fewer noisy actions. M2 correlated most highly with people's data at relatively low values of γ. This allowed M2 to integrate evidence for goals from past actions, inferring goal changes only rarely, when there was sufficiently strong support for them. M3 correlated most highly with people's data at intermediate κ values. This allowed M3 to infer subgoals when necessary to explain large deviations from the shortest path to the final goal, but to avoid positing them unnecessarily for paths that had a simpler explanation.

Third, was the pattern of errors of each model class informative? Did any particular class of trials drive the differences in correlation between models? Given that some of the differences between models, while statistically significant, were small, do the differences in model fits reflect genuine differences in the ability of these models to capture people's mental representations of agents' goals? To address these questions, we performed a targeted analysis of a class of experimental trials that exposed key differences in how the predictions of M1, M2, M3 and H depended on recent and past information from agents' paths. The targeted analysis focused on data from all of the Gap obstacle conditions, at judgment points 10 and 11. Example conditions are shown in Fig. 6(a), with judgment points 10 and 11 circled. In Fig. 5, the black points represent ratings from the targeted analysis conditions, which account for all of the outliers of H and most of the outliers of M1 and M3.

In the targeted conditions, the agent's recent actions at judgment point 10 were always ambiguous between goals A and B. However, the entire path provided evidence for either A or B depending on whether it went around or through the obstacle. In conditions 1 and 2 of Fig. 6, participants used the information from the agent's early movements, rating A nearly twice as likely as B because the agent had gone around, not through the obstacle. M1, M2 and M3 captured this pattern by integrating evidence over the entire path, while H did not, rating A and B nearly equally because the most recent movements were ambiguous between the two goals. At judgment point 11, the agent either took a further step that remained ambiguous between goals A and B, or moved unambiguously toward a particular goal. In condition 1 of Fig. 6, the agent's action at judgment point 11 remained ambiguous between A and B, but participants continued to favor A due to the path history. Again, M1, M2 and M3 predicted this pattern, while H did not. In condition 2 of Fig. 6, the agent's action at judgment point 11 strongly supported B – a complete reversal from previous evidence. Now, participants inferred that B was the goal, weighing the strong recent evidence over the accumulated past evidence for A. M2 and H matched this pattern, M3 captured it only weakly, and M1 did not, still favoring A based on all the past evidence against B.

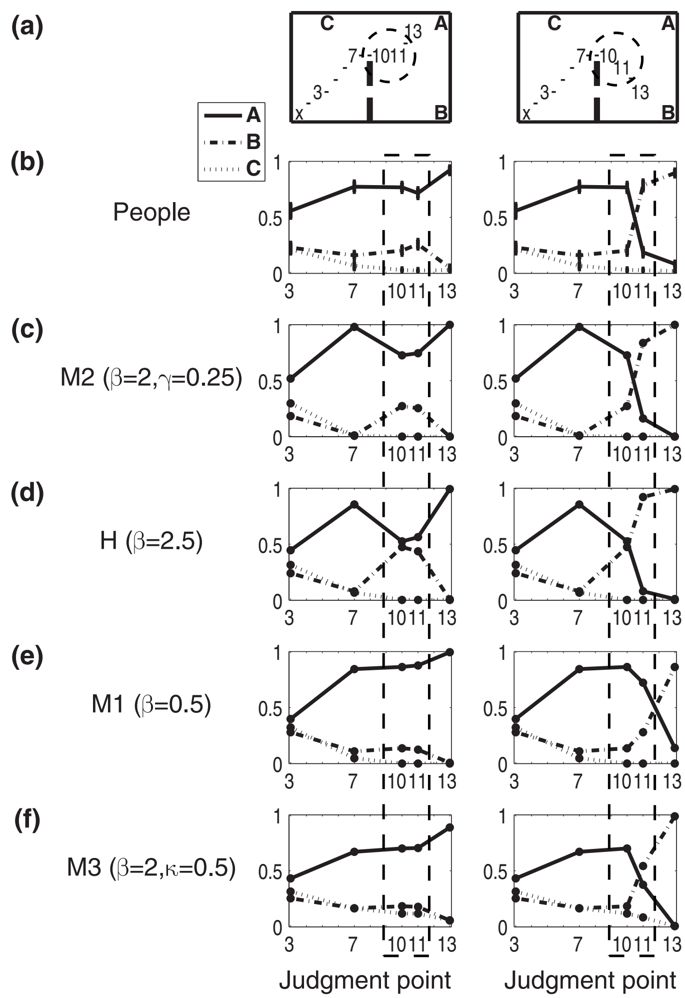

Figure 6. Targeted analysis of Experiment 1. **(a)** Example conditions from our targeted analysis. **(b)** Participants' ratings from the above conditions. **(c)** Predictions of M2 ($\beta = 2.0$,$\gamma = 0.25$). **(d)** Predictions of H ($\beta = 2.5$). **(e)** Predictions of M1 ($\beta = 0.5$). **(f)** Predictions of M3 ($\beta = 2.0, \kappa = 0.5$).

Discussion

Experiment 1 showed that inverse planning models based on simple structured goal priors can predict people's online goal inferences very accurately in the maze-world domain. M2 predicted participants' judgments most accurately overall, correlating significantly higher with people's judgments than the alternative models we considered. Although M2 is a more complex model than M1 or H, with an additional parameter representing the probability of goal switching, our analysis accounted for this, using model selection techniques to measure the generalization performance of each model while preventing overfitting and controlling for model complexity.

We probed people's assumptions about goal switching with two additional analyses. First, in a targeted analysis of trials where the agent's recent actions were locally ambiguous between goals, but globally unambiguous, M2 predicted the pattern of subjects' judgments more accurately than alternative models. These targeted trials comprised nearly all of the largest outliers for the alternative models, which suggested that M2's representation of the probability of goal switching was necessary to explain people's goal inferences in this context. Second, we found that the best-fitting parameters values for the goal switching prior for M2 were low, but nonzero, consistent with the notion that the goals of an intentional agent may switch, but tend to persist across time.

In sum, M2 accounted best for participants' data from Experiment 1. However, the lower correlations of M3 and H were mainly driven by a small number of experimental trials; for the majority of trials, M1, M2, M3 and H all made similar predictions, belying the essential differences in the way these models parsed actions. Our next experiment looked for more qualitative differences between these models in a new task, where models that assumed an agent's goal could change over time (M2 and H) made qualitatively different predictions from models that were constrained to infer a single goal (M1 and M3).

Experiment 2

Experiment 2 presented a new task based on retrospective goal inference. Using stimuli derived from Experiment 1, we showed subjects only paths ending at the longest judgment points from Experiment 1 and asked them to make retrospective inferences about agents' goals at earlier judgment points in the action sequence. Models based on static goals (M1 and M3) and models based on changing goals (M2 and H) made qualitatively different predictions in this task. M1 and M3 were constrained to parse an agent's actions in terms of a constant, global goal (consisting of a single goal for M1, and possibly containing a subgoal for M3). This meant that the goal inferences made by M1 and M3 did not change between judgment points within a particular condition, because the same full path was displayed in each trial. M2 and H parsed actions in terms of a sequence of goals. M2 was biased to infer goals that changed infrequently, while H assumed no dependence or consistency between an agent's goals at each step. Comparing how accurately M1 and M3 predicted people's retrospective goal inferences with the accuracy of M2 and H provided further evidence for the concept of changing goals in human action understanding.

Method

Participants.

Participants were 16 members of the MIT subject pool, 10 female, 6 male.

Stimuli.

Stimuli for Experiment 2 were derived from Experiment 1 stimuli as follows. For each trial of Experiment 1, an Experiment 2 trial first displayed an animation of an agent's entire path, up to the longest judgment point from that condition. Then, an intermediate judgment point along the path, taken from that condition of Experiment 1, was marked with a red "+". Subjects then reported their retrospective goal inferences: how likely each goal was when the agent was at the marked judgment point. Fig. 2(b) shows the appearance of Experiment 2 stimuli alongside corresponding Experiment 1 stimuli.

Design.

Experiment 2 used the same set of conditions as Experiment 1, all represented in Fig. 3. Because of this, subjects' ratings could be compared between experiments to assess the different effects of online and retrospective goal inference tasks. In each condition, all but the longest judgment point trials (which provided the endpoint for each displayed path) and judgment points of 10 (excluded to reduce the total number of stimuli) were taken from Experiment 1. In Experiment 2, each condition displayed a unique combination of goal configuration, obstacle shape and complete path, so there was no overlap between judgment point trials of different conditions as there was in Experiment 1. Experiment 2 had 95 stimuli in total.

Procedure.

Subjects were given a modified version of the cover story from Experiment 1. In Experiment 2, subjects were again told that they would view videos of members of an intelligent alien species collected by scientists, but this time, they were told: "the scientists are not sure how the aliens decide where to go, but the aliens generally move toward goals in the environment that are labeled for you with capital letters." They were then told that after the video of each alien's movement, a point along the path would be marked, and they would rate which goal the alien "had in mind" at that marked point. The stimulus ordering and rating procedure was similar to Experiment 1, except that now subjects rated the likelihood of different goals at earlier points in agents' action sequences, rather than at the end of their action sequences as in Experiment 1.

Modeling.

We modeled Experiment 2 in the same manner as Experiment 1, except now model predictions were given by retrospective versions of Equation 1 (specific versions for M1, M2, M3 and H are provided in the appendix).

Results

Our analysis of Experiment 2 paralleled our analysis of Experiment 1, combining qualitative and quantitative comparisons of how accurately M1, M2, M3 and H predicted people's retrospective goal inferences. As we will argue below, models that allow goal changes, such as M2, best explained people's judgments in Experiment 2. Fig. 7 shows qualitative comparisons between participants' goal inferences and the predictions of M2. There is a direct correspondence between the conditions shown in Fig. 7 and the conditions in Fig. 4 from Experiment 1, which allows the effects of the online task of Experiment 1 versus the retrospective task of Experiment 2 to be assessed directly. Fig. 7(a) shows retrospective versions of the corresponding conditions in Fig. 4(a). Each condition in Fig. 7(a) differs from adjacent conditions by one stimulus feature, providing examples of the effect of changing the goal configuration, the obstacle shape and the agent's path.

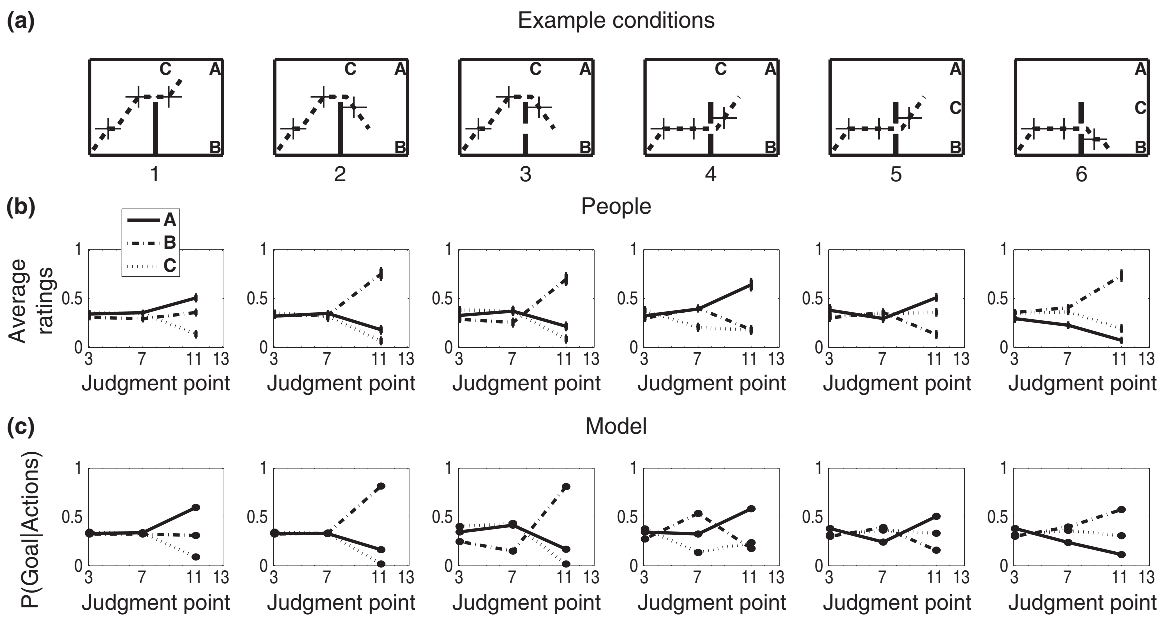

Figure 7. Example conditions, data and model predictions from Experiment 2. **(a)** Stimuli illustrating a range of conditions from the experiment. These stimuli directly correspond to the Experiment 1 stimuli in Fig. 4(a). Dashed lines correspond to the movement subjects saw prior to rating the likelihood of each goal at each judgment point, which are marked by black +'s. **(b)** Average subject ratings with standard error bars for the above stimuli. **(c)** Predictions of inverse planning model M2 with parameters $\beta = 0.5$, $\gamma = 0.65$.

Fig. 7(b) shows participants' average ratings with standard error bars for each of these conditions. Many of the same patterns of reasoning occurred in our Experiment 2 data as in Experiment 1. Participants' inferences again tended to become more certain at judgment points closer to the end of each path. In general, however, people's ratings from Experiment 2 reflected greater uncertainty than in Experiment 1. In the task of Experiment 2, people's inferences eventually approached the goal that was suggested at the end of the path. These trends were predicted by M2 and H, but not M1 or M3, which were constrained to make the same goal inference at each judgment point within a particular condition.

Fig. 7(c) shows the predictions of M2 with the best-fitting parameter values. Comparing Fig. 7(b) and (c), the model predicted people's judgments very accurately for the examples shown. M2 explained people's retrospective judgments based on inferring the sequence of goals that best explained the agent's path. The agent's actions near the end of each path always strongly suggested a particular goal, but earlier actions could be ambiguous, or indicate a different goal. M2 integrated this past and future information to infer the goal at each judgment point that was most consistent with an agent that took approximately rational paths toward its current goal, but could occasionally change its goal.

We quantitatively compared how accurately M1, M2, M3 and H predicted participants' judgments with a bootstrap cross-validated (BSCV) correlational analysis. The average correlations of each model with people's data are shown in Table 2. M2 predicted people's judgments most accurately, correlating significantly higher than M1 ($p_{BSCV} < 0.0001$), M3 ($p_{BSCV} < 0.0001$) and H ($p_{BSCV} = 0.0168$). H performed second best, correlating significantly higher than M1 ($p_{BSCV} < 0.0001$) and M3 ($p_{BSCV} < 0.0001$). Finally, M1 and M3 both performed poorly, and the slightly higher correlation of M3 than M1 was not significant ($p_{BSCV} = 0.44$).

	M1	M2	M3	H
$\langle r \rangle$	0.57 (0.016)	0.95 (0.0077)	0.58 (0.019)	0.91 (0.0069)

Table 2: Bootstrap cross-validated correlation of inverse planning models and a simple heuristic alternative with people's data from Experiment 2. Numbers in parentheses indicate standard deviation.

Scatter plots of the correspondence between participants' judgments and M1, M2, M3 and H using the best-fitting parameter values are shown in Fig. 8 to illustrate the pattern of errors for each model. M2 predicted people's judgments most accurately, and in Fig. 8, has no clear outliers from the line-of-best-fit. H predicted people's judgments less accurately than M2, and M1 and M3 performed worst and have the most outliers in Fig. 8.

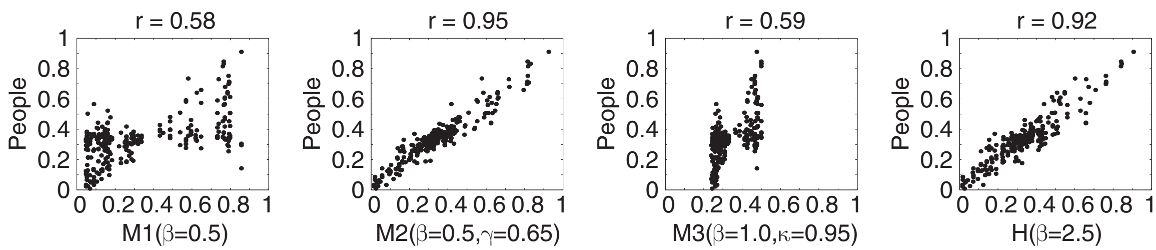

Figure 8. Scatter plots of model predictions using best-fitting parameter values (X-axes) versus people's retrospective goal inferences (Y-axes) for all Experiment 2 stimuli.

To assess the dependence of these results on the parameter values, we tested how accurately M1, M2, M3 and H predicted people's goal inferences under a range of parameter settings. Plots of the correlation of M1, M2, M3 and H with participants' data across all tested values of β, γ and κ are shown in the appendix. In Experiment 2, all models fit subjects' data best with lower β values than in Experiment 1, indicating that subjects assumed a noisier agent in Experiment 2. For M2, relatively high γ values yielded the best correlation, suggesting that subjects were more willing to infer goal changes in Experiment 2 than in Experiment 1. At high γ values, M2 predicted that people's ratings would depend strongly on the information provided by agents' movements close to the retrospective judgment point. H correlated best with people's judgments with slightly higher β values than M2, because of the additional flexibility provided by assuming that the agent's goal could change arbitrarily at each timestep. Both M1 and M3 performed poorly because they made the same predictions at all judgment point trials within each condition, while participants' goal inferences clearly varied at different steps of agents' paths.

Discussion

Experiment 2 showed that inverse planning models based on priors that allow changing goals can predict people's retrospective goal inferences very accurately in the maze-world domain. In an analysis that controlled for model complexity using bootstrap cross-validation, M2 correlated significantly higher with participants' judgments than models that assumed static goals (*e.g.* M1 and M3). Although the predictions of M2 and H were qualitatively similar, M2 also correlated significantly higher with participants' judgments than H.

A further analysis probed the parameter values for each model that best explained participants' data. In Experiment 2, the highest correlations for all models occurred at lower β values than in Experiment 1, indicating that people assumed "noisier" agents in Experiment 2. This may have been due to several factors. First, our instructions told subjects that "the scientists are not sure how the aliens decide where to go", which may have led them to assume noisier agents. Second, the retrospective task of Experiment 2 may have been more challenging than the online task of Experiment 1, making subjects less sure of their judgments. The practical consequence of this was that evidence from agents' movements affected subjects' inferences less strongly in Experiment 2 than in Experiment 1. Higher values of γ also yielded better correlations with subjects' judgments in Experiment 2 (higher κ values technically yielded better correlations as well, but this difference was negligible). Because the retrospective goal inference task presented a context where people inferred changing goals, it may have also biased subjects toward responses consistent with higher values of γ.

The results of Experiment 2 suggested that the concept of changing goals is crucial for retrospective goal inferences. Representing goal changes allows inferences of past goals and current goals to differ with sufficient evidence, providing an adaptive forgetting factor that can weigh recent evidence more strongly than past evidence. Goal switching is just one way to extend the concept of a single, static goal, but by no means the only way. The representation of complex goals in M3 is surely also important in some contexts. Experiment 3 explored such a context.

Experiment 3

Experiment 3 probed how a range of different behaviors and situations might lead people to infer subgoals. Any observed behavior can be explained by an infinite number of goals, and more complex goals can always fit an observed behavior better than simpler goals, by positing a sequence of subgoals that follows the observed path arbitrarily closely. For example, if a person usually takes a fairly direct route home from work, with only occasional small deviations or apparent detours, they probably have a simple, constant goal: to get home. However, they could also sometimes have subgoals that cause the deviations: maybe one day they wanted to stop off at the market on the way home, on another day they wanted to stop off at the post office, and so on. There is a danger of "overfitting" in positing these subgoals, as any random deviation could in principle be explained this way.

Stronger evidence that an agent truly has one or more subgoals would come from observing consistent deviations from the shortest path home: if a person regularly appears to go a few blocks out of their way at the same spot and time of day, it suggests an intentional action, such as stopping for groceries on the way home from work. Even greater evidence for subgoals can come from observing consistent action sequences under different starting conditions, particularly when they lead to large apparent detours from the shortest path to the final goal. For instance, if a person occasionally heads home not from work but from different points (a doctor's office, an offsite meeting), and still passes through the same apparent subgoal locations no matter how far away, it is a strong sign that they specifically intend to visit those locations.

These patterns of evidence are related to the notion of equifinality (Heider, 1958), a classic cue for intentional attribution from social psychology, based on identifying invariant effects of actions across multiple situations. Our framework explains inferences about whether an agent is pursuing a simple goal or a complex goal with subgoals through a version of the Bayesian Occam's razor (Jefferys & Berger, 1992). In general, Bayesian inference naturally trades off simplicity and fit to the data in evaluating hypotheses. In the context of action understanding with probabilistic planning models, Bayesian inference weighs the greater likelihood that comes from explaining more detailed variation in an observed trajectory against the lower prior probability that is assigned to more complex goals, with more subgoals.

To test whether people's judgments were consistent with these principles, Experiment 3 used a new task involving prediction of an agent's future behavior, given examples of its behavior in the past. This experiment used simple maze-world versions of the scenarios described above. Our stimuli presented one or more trajectories in which the agent seemed to change direction at a single midpoint location along the way to a final goal, suggesting the possibility of a complex goal with a subgoal. We assessed whether subjects inferred a subgoal by asking them to predict the agent's hypothetical trajectory starting from a different initial location; if they inferred a subgoal, they would predict a different, more indirect path than if they inferred only a simple final goal. Different conditions varied the amount of evidence for complex goals. In some cases, intended to suggest a simple goal to observers, the change in direction could be naturally explained by environmental factors (avoiding an obstacle) or as a correction from a small random path deviation. In other cases, observers saw multiple trajectories starting from different initial positions, all with the same intermediate switch point, which should strongly suggest a subgoal at that location. We compared people's judgments in these different conditions with those of our Bayesian inverse planning models, using the models' ability to predict future action sequences consistent with the goals inferred from earlier observed action sequences (Equation 2).

Method

Participants.

Participants were 23 members of the MIT subject pool, 14 female, 9 male.

Stimuli.

Experiment 3 used a maze-world stimulus paradigm similar to Experiments 1 and 2. Fig. 9 shows all stimuli from Experiment 3. Each stimulus displayed a complete action sequence, ending with the agent achieving its goal. The environment had only one visible goal, marked by a small orange triangle, and the size of the environment was 17 squares wide by 8 squares high.

Design.

Experiment 3 used a $2 \times 2 \times 2$ factorial design. Each row of Fig. 9 shows a different condition of Experiment 3, and conditions are divided into groups A, B, C, and D. The first factor varied the directness of the agent's paths to the goal. Conditions A and C displayed direct example paths to the goal, while conditions B and D displayed indirect example paths. The second factor varied the presence or absence of an obstacle in the environment. Conditions A and B displayed environments with a wall-like obstacle extending up from the bottom edge, and conditions C and D displayed environments without the obstacle. The third factor varied the location of the subgoal relative to the location of the marked goal, which affected the length of the deviation from the direct path when the agent went through the subgoal. In the "Far" subgoal conditions, the subgoal was farther from the marked goal, given by the grid square directly above the obstacle, resulting in a larger deviation from the most direct path. In the "Near" subgoal conditions, the subgoal was nearer to the marked goal, corresponding to a point between the obstacle and the marked goal, resulting in a smaller deviation from the most direct path.

Each condition presented 4 trials, which varied the number of paths subjects saw before predicting the agent's action in a new situation. Each trial had two phases. In the example phase of each trial, a new, animated example path was presented to subjects, following the order shown in Fig. 9. The example paths varied the agent's starting point in the environment, but always ended with the agent reaching the same final point. The example paths in each condition had additional structure beyond the factorial design: conditions A, B and D displayed example paths consistent with a complex goal, while condition C displayed paths that were not consistent with any single complex goal (*i.e.* there was no common subgoal that all example paths passed through). Example paths from corresponding trials of conditions A, C and D all started from the same point, while paths from corresponding trials of condition B started from these points reflected to the opposite side of the obstacle.

Next, in the response phase, two different hypothetical paths that the agent could take if starting from a different location in the environment were displayed. All response paths started and ended at the same point, and used the same environments and goal locations as the example paths in their respective conditions. Each pair of response paths featured one direct path to the goal and one indirect path to the goal, with the indirect path passing through the subgoal from its condition. Conditions with the same subgoal location all used the same response paths.

Procedure.

Before each condition, subjects were told verbally that they would be watching a series of videos of intelligent aliens moving in their natural environment, and that each video would show the same alien moving in the same environment, with walls through which aliens could not move. They were told that after each video, they would see two possible movements the alien might take when starting from a different location, and that

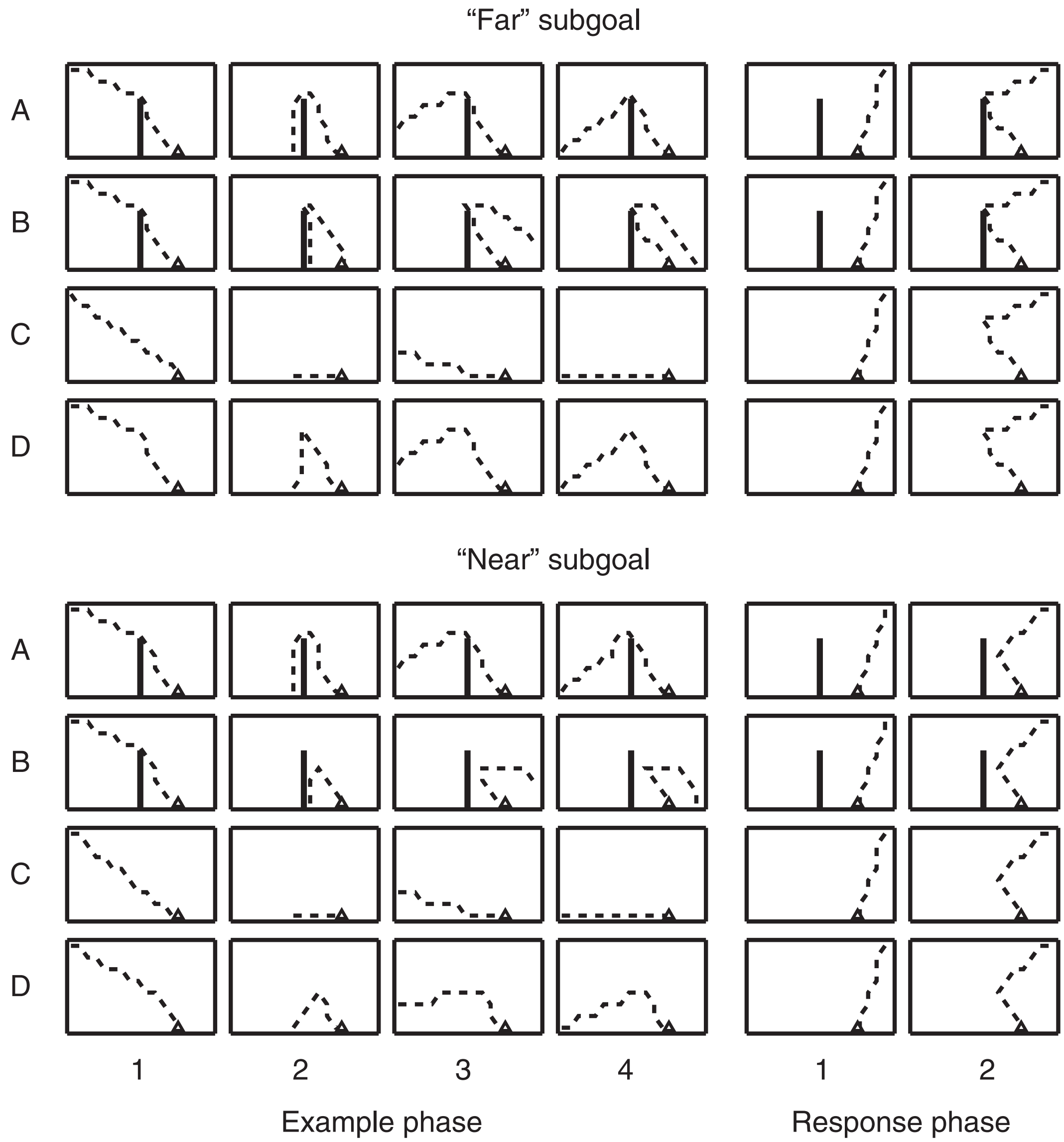

Figure 9. All stimuli from Experiment 3. The experiment followed a 2 × 2 × 2 design, which varied the directness of the agent's path, the presence or absence of an obstacle, and the location of potential subgoals. Each condition had 4 trials, which each presented a different example path, and then asked subjects to predict which of 2 response paths was more likely given all previously displayed example paths from that condition.

they would rate how likely each movement would be given their observations of previous movements.

The first trial of each condition presented an animation of the first example path. Subjects then rated the relative likelihood of the two response paths on a 9-point scale from 1: "Definitely path 1", to 5: "Paths 1 and 2 are equally likely", to 9: "Definitely path 2". In subsequent trials of each condition, animations of the next example path were presented, and all previous example paths remained onscreen to aid recall. Subjects were instructed to base their judgments on the evidence from all the agent's example paths in that condition.

Stimuli were shown from an overhead perspective, with an animated schematic trace of the agent's path as it moved through the environment. After each example path was displayed, response paths were presented simultaneously, and the side of the screen on which each response path was displayed was randomized in different conditions. Stimulus conditions were presented to subjects in pseudo-random order. To minimize presentation order effects, environments and paths were reflected and rotated in randomized increments of 180 degrees between conditions to reduce the impression of repeated paths. To further minimize order effects between conditions, after each condition, subjects were told verbally that they would observe a new alien, of a different color than the previous alien(s). They were told that this alien might be different than the previous alien(s), so they should not use information about the previous alien(s) when making judgments about the new alien.

Modeling.

Model predictions take the form of probability distributions over agents' future actions, conditioned on their previously observed actions, given by Equation 2. To compare model predictions with subjects' judgments, we computed the log posterior odds ratio of the two response paths in each condition, conditioned on the previously observed example paths. We then mapped the z-scores of the log posterior odds ratio through the sigmoidal standard normal cumulative density function, and computed correlation coefficients between these values and participants' average ratings. We discuss this analysis further in the appendix. M3 assumed that the agent had either a simple goal or a complex goal with one subgoal, and that both goals and subgoals could be visible or invisible, and could correspond to any grid square in the environment.

Results and Discussion

Our analysis of Experiment 3 compared subjects' judgments with the predictions of our models to test whether people's goal-based generalization could be explained by Bayesian inverse planning. We will focus on the predictions of M3, because it was the only model that could represent and infer the kinds of complex goals presented in Experiment 3. M1 performed poorly in this task, with a bootstrap cross-validated correlation coefficient of $\langle r \rangle = -0.03$. M2 and H did not naturally allow goal-based generalization, and an extension of them to do so performed poorly as well, with a bootstrap cross-validated correlation coefficient of $\langle r \rangle = 0.54$ for both models. We analyze these models further in the appendix, but it is easy to characterize their problems in intuitive terms. M1 is too inflexible, always predicting that an agent will take the most direct path to its final goal. M2 and H, in contrast, are too flexible. They explain apparent subgoals as arbitrary changes in the agent's final goal, which leads to predictions of future arbitrary goal changes that cause noisy, unstable behavior.

Fig. 10 compares participants' average ratings from all conditions of Experiment 3 with the predictions of M3 using the best-fitting parameter values. M3 predicted people's judgments very accurately overall, and also predicted the qualitative effects of varying each stimulus factor. M3 correctly predicted that subjects would infer a subgoal in conditions B and D, but not in conditions A and C. M3 also captured the rate at which evidence accumulated for or against the subgoal in each condition, fitting each of the learning curves across trials very closely.

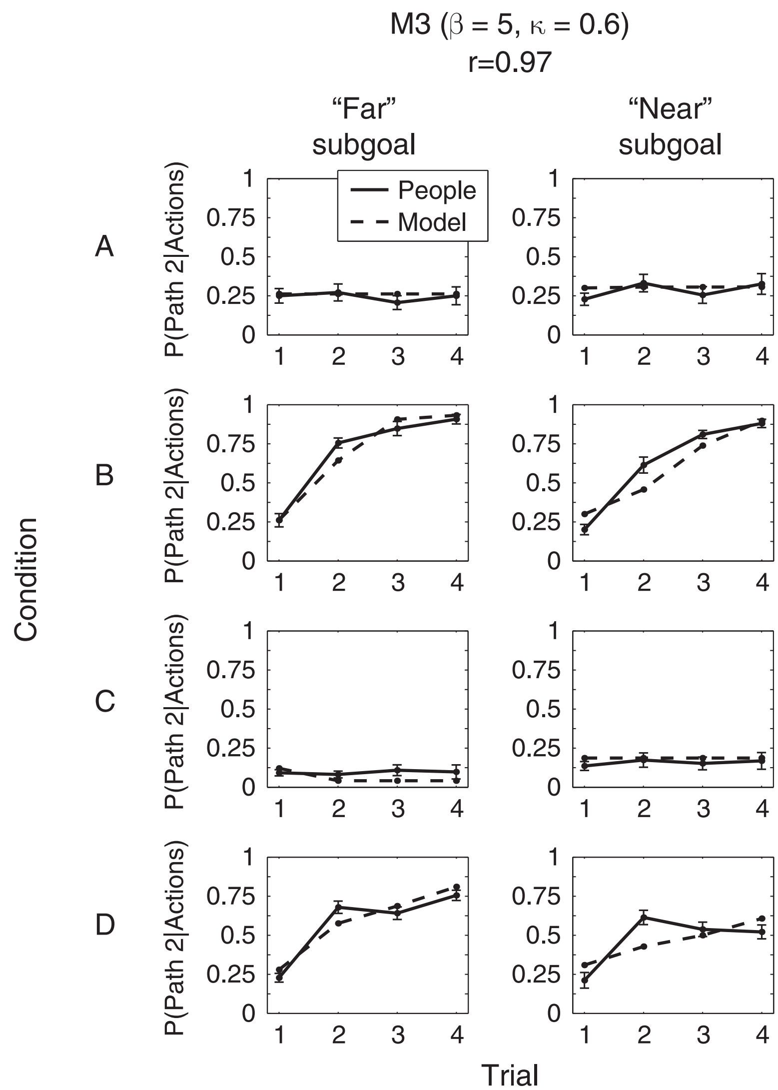

Figure 10. Subjects versus M3 with the best-fitting parameters for all stimuli from Experiment 3. Ratings correspond directly to conditions from Fig. 9.

Two intuitive principles underlie these predictions. First, a sequence of actions provides more evidence for a subgoal interpretation to the extent that it embodies a larger deviation from the shortest path linking the agent's starting point to the final goal. The larger the deviation, the more of the agent's actions would have to be attributed arbitrarily to chance under a simple goal interpretation – lowering the probability of that interpretation and raising the probability of a subgoal. This explains why the paths in conditions B and D suggested a subgoal while the condition A paths did not. It also explains why condition B provided stronger evidence for a subgoal than condition D, and why the "Far" subgoal conditions provided stronger evidence than the "Near" subgoal conditions. Second, because the agent's goal was assumed to be constant across trajectories within a single condition, evidence for a subgoal accumulated when the same midpoint location appeared to be a subgoal in multiple trajectories starting from different initial conditions. This explains why subgoal inferences became stronger with more evidence in conditions B and D, and why they were even weaker in condition C than condition A – because A paths were consistent with a single candidate subgoal while C paths did not even pass through a common midpoint location.

The quantitative fit of M3 to subjects' data was also very accurate. The bootstrap cross-validated correlation coefficient of M3 with subjects' judgments was $\langle r \rangle = 0.96$, which was significantly higher than the correlation of the other models ($p_{BSCV} < 0.0001$). To assess the dependence of the model predictions on the parameter values, we compared subjects' judgments to model predictions under a range of different parameter settings. Plots of the correlation of our models with participants' data across all tested parameter values are shown in the appendix. M3 correlated very highly with subjects' judgments for a wide range of parameter values, with $\beta \geq 2.0$ and $\kappa \leq 0.95$. The best-fitting correlation for M3 occurred with $\beta = 5.0$ and $\kappa = 0.6$, with a correlation coefficient of $r = 0.97$. The best-fitting β value was higher than in previous experiments because the paths in Experiment 3 were less noisy, and M3 could assume a more deterministic agent.

In sum, Experiment 3 showed how inverse planning models that represent subgoals can capture human action understanding in the maze-world domain, and can generalize from previous observations to predict novel action sequences by inferring invariant goals. Our subgoal model M3 predicted people's judgments with high accuracy, explaining the strength of subgoal inferences based on two intuitive stimulus factors: the number of independent action sequences consistent with a putative subgoal, and the length of the deviation in these paths relative to the shortest paths from each initial state to the final goal.

General Discussion

We presented a computational framework for modeling human action understanding, and some of the first combined experimental and computational studies of adult goal inference and action prediction. Our studies made three main contributions. First, we presented strong evidence that human action understanding can be formalized as Bayesian inverse planning in Markov decision problems. In quantitative terms, our models correlated highly with people's judgments across multiple conditions and experiments. They also provided insights into several qualitative phenomena of goal inference, such as the effect of alternative available routes in Experiments 1 and 2, or the dependence of subgoal inferences on the length of deviations from shortest paths in Experiment 3. Our experiments examined only one domain of spatial navigation in simple mazes, but this task shares deep similarities with more complex, naturalistic tasks. Likewise, our framework of inverse planning in Markov decision problems extends to much richer settings. MDPs can be generalized to partially observable environments (Kaelbling et al., 1998) and multi-agent situations (Littman, 1994; Filar & Vrieze, 1997), and these generalizations can model domains and tasks that go well beyond a single agent moving in a simple two-dimensional maze, such as motor action (Todorov, 2004; Körding, 1997), games and other strategic interactions (Littman, 1994; Filar & Vrieze, 1997), and a range of cooperative or communicative activities (Littman, 1994). In recent work, we have used these ideas to build inverse planning models of how people infer "social goals" in multi-agent interactions, such as whether one agent is chasing or fleeing from another (Baker et al., 2008); see Yoshida et al. (2008) for a similar approach.

Second, our experiments provided evidence that inverse planning models can predict people's judgments about an agent's goals more accurately than a simple heuristic alternative that looks only at the agent's current heading. Similar heuristics have been proposed to explain how people categorize movement as intentional or animate in stimuli similar to our own (Blythe et al., 1999; Zacks, 2004; Barrett et al., 2005). We showed how to formulate this heuristic as a limiting case of inverse planning, providing insight into the situations under which it approximates our ideal observer models of inverse planning. In these cases, in particular where an agent appears to follow a single shortest path to a fixed goal, the heuristic indeed predicted people's judgments accurately. However, in cases where the approximation breaks down and the heuristic's predictions diverge from our more general inverse planning models, the latter gave a much better fit to people's goal inferences.

Third, we showed how our framework can be used to distinguish different hypotheses about the contents of people's mental representations of agents' goals, and provided evidence for the importance of different kinds of goal representations in different contexts. Experiment 1 showed that in online goal inference, goal representations must be more flexible than just single fixed locations (M1); representations based on changing goals (M2) or subgoals (M3) fit people's judgments significantly better, with a small advantage for the former. Experiment 2 provided definitive evidence for the use of representations of changing goals (M2) in a retrospective judgment task. Experiment 3 showed that given sufficient evidence – multiple trajectories starting from different positions but consistent with the same final goal and subgoal locations – people would engage more complex subgoal-based representations.

Our models rest on a number of general assumptions about the background knowledge underlying human action understanding, as well as specific assumptions about how subjects approach our laboratory tasks. The rest of this discussion attempts to lay out these assumptions and sketches some of our ongoing work aimed at relaxing them or explaining them in a more satisfying way.

On each individual trial of the experiments, our subjects and our models observed the same data, but the models were also given certain basic aspects of the MDP setup, such as the agent's cost function and the probabilistic dynamics by which the agent moves through the environment. Since subjects were not explicitly told this information, how is it reasonable to model their judgments as if they knew it? People certainly come to our tasks – and any action understanding task – with rich background knowledge about intentional action that goes far beyond what we have assumed here. This background knowledge plausibly includes the assumptions made by our models: (1) that the agent's cost function depends on distance traveled to the ultimate goal, (2) that the agent probabilistically selects actions as a function of their expected value, and (3) that the agent's actions yield the intended state transitions. Indeed, these assumptions are at the heart of previous qualitative descriptions of human action understanding, like the "teleological stance" (Gergely et al., 1995) and simple versions of the "intentional stance" (Dennett, 1987). Nevertheless, future research should explicitly test these assumptions.

Another crucial piece of background knowledge assumed by our models, and conveyed to our participants in the instructions, is that the entity under observation should be treated as a intentional agent. In some real world contexts, by contrast, people must infer whether an observed entity is in fact a rational or intentional agent. These inferences are likely a function of both the entity's observed actions in the context, and the observer's prior expectations based on how the entity appears or is described. Even young infants can apply a teleological stance to a novel entity, without being explicitly told it is appropriate, based either on the agents' actions (Csibra et al., 1999) or its appearance (Johnson et al., 1998; Guajardo & Woodward, 2004; Saxe et al., 2005).

In ongoing work, we are modeling the inference that an entity observed in motion is an intentional agent, using a hierarchical Bayesian model (HBM) (Good, 1980; A. Gelman et al., 2003; Tenenbaum et al., 2006; Kemp et al., 2007). An HBM represents data at multiple levels of abstraction. It allows us to consider probabilistic models of both intentional action and other kinds of motion, due to inanimate objects or artifacts, or animate but non-intentional beings. Classifying an agent as an intentional actor can be captured by an HBM with two levels, with the higher level selecting among alternative models that could explain some observed motion (intentional, inanimate, animate), and the lower level selecting among specific explanations generated by these models (*e.g.*, a specific goal interpretation that explains the particular trajectory that a putatively intentional agent appears to follow). Given observations of how an entity moves in some context, these models can simultaneously make inferences over a hypothesis space of both intentional and non-intentional models, evaluating each model, as well as the specific hypothesis within the classes generated by these models, based on how well they explain the observed motion.

Hierarchical Bayesian models may also provide the solution to a puzzle in our current data. In our experiments, two different models of goal-directed action provided the best fit to participants' judgments: in Experiments 1 and 2, M2 (which allowed for agents to change their goals) provided the best fit, while in Experiment 3, the best fit was with M3 (allowing subgoals within a single action sequence). Did our participants choose different kinds of goal representations for the actions presented in these experiments, and if so, how did they make this choice? We hypothesized that participants actually made a rational inference based on the stimuli they observed, selecting not only the specific goal that best explained the agent's action on any one trial, but also selecting the class of goal representations that best explained all of the agents' actions observed across multiple trials. Selecting which class of goal representations to use in a particular context can also be captured by an HBM with two levels, with the higher level selecting a hypothesis space of goals (M1, M2, M3 or H), and the lower level selecting a specific goal from the hypothesis space given by the higher level. This goal then generates a sequence of actions, conditioned on the agent's environment.

We make this HBM analysis precise in the appendix. For each model (M1, M2, M3 or H) and each experiment (1, 2 or 3), we compute $\log P(\mathsf{Stimuli}|\mathsf{Model})$, a measure of how well that model explains all the stimuli observed in that experiment as a whole. The results of this analysis are shown in Table 3. In each case, the model with the highest marginal likelihood (shown in bold) is the model that correlated most highly with people's judgments in that experiment. Thus, the goal representation that people appear to use in each experiment is also the one that an ideal learner would see as the best explanation for agents' observed behavior in that experiment.

	log P(Stimuli|Model)			
	M1	M2	M3	H
Experiment 1	-773.6	**-641.9**	-651.6	-778.5
Experiment 2	-1118.5	**-832.4**	-860.1	-1068.1
Experiment 3	-457.2	-294.3	**-236.5**	-391.5

Table 3: Log marginal likelihood of models and heuristic given all Experimental stimuli. Higher log-likelihood indicates a better fit, and the highest value for each experiment is shown in bold.

The content of people's representations of agents' mental states is surely much more complex than the models we consider here. Important directions for future research are to apply our models in more complex environments, and to continue exploring the space of complex goal structures. A potential framework for extending our models to structured, probabilistic environments is given by factored MDPs (Guestrin et al., 2003). This framework represents the environment as a dynamic Bayesian network, and allows efficient algorithms for planning over much larger and more complex state spaces than those feasible in standard MDPs. A possible representation language for hierarchical goal structures is provided by hierarchical MDPs (Dietterich, 2000; Parr & Russell, 1998), which provide a natural hypothesis space for more complex goals, and can represent the space of goals we consider here. With an appropriate prior on the space of models of goals, it should be possible to learn new goal structures from data using hierarchical Bayesian models like those we sketched above.

Another important direction for future research will be to extend our framework to modeling people's reasoning about agent's beliefs. This can be formulated as inverse planning using partially observable MDPs (POMDPs) (Kaelbling et al., 1998) to handle joint inferences about agents' beliefs and goals. A sketch of this extension was described in Fig. 1(c). Here we focused on the special case of Fig. 1(b) because it has been the target of extensive empirical work with infants and adults, and also because it is computationally much simpler to implement. Optimal planning in POMDPs is computationally intractable, so we would require some approximate algorithm to generate expectations about agents' beliefs and goal-dependent behaviors. In ongoing work we are beginning to explore the more general POMDP case, along with various approximate inference schemes based on cognitively plausible approximations to ideal optimal planning.

Do people really interpret others' actions using a causal model structured around an explicit assumption of rationality, as our models do, or do they instead simulate others' planning processes using their own planning mechanisms – in effect building in an implicit assumption of rationality to the extent that their own planning mechanisms are rational? The debate between "theory-based" and "simulation-based" accounts of action understanding has generated much attention in the recent literature (Gopnik & Meltzoff, 1997; Goldman, 2006) and on first glance our work appears most consistent with the "theory-based" approach. Formalizing an intuitive theory of mind was in fact one of our original motivations, and elsewhere one of us has argued for theory-based accounts against simulation accounts (Saxe, 2005). However, the models we propose here could be sensibly interpreted under either account. On a theory-based interpretation, inverse planning consists of inverting a causal theory of rational action to arrive at a set of goals that could have generated the observed behavior, and inferring individual goals based on prior knowledge of the kinds of goals the observed agent prefers. On a simulation account, goal inference is performed by inverting one's own planning process – the planning mechanism used in model-based reinforcement learning – to infer the goals most likely to have generated another agent's observed behavior. In future work, we hope to be able to distinguish these accounts by testing whether observers' interpretations of other agents' behavior in a particular context can be predicted by inverting policies measured from how the observers themselves plan actions in the same context.

Conclusion

Formal models of "ideal observers" or "ideal inference agents" have long played an important role in the study of core cognitive capacities, such as visual perception, memory retrieval, or language processing and acquisition (Liu et al., 1995; Weiss et al., 2002; Shiffrin & Steyvers, 1997; Anderson, 1990; Hale, 2001; Gold, 1967). These models allow us to assess how well and in what ways people's mental representations of the world correspond to reality, by seeing how close people's inferences come to the best possible inferences under different assumptions about how the relevant stimuli might be represented and processed. Human percepts or inferences in some domains, particularly lower-level functions such as visual motion or surface perception, often come remarkably close to the ideal limit.

Here we have taken a first step toward extending the ideal-observer approach to a higher-level perceptual task of action understanding, specifically goal inference. Human action is notoriously complex and difficult to predict, perhaps even irrational in many situations (Kahneman et al., 1982; Hernstein & Prelec, 1991); it is certainly more complex than the moving-dot or slanted-surface stimuli for which ideal observer analyses are better known in vision. It is perhaps surprising then that people's judgments about agents' goals appear so consistent with an ideal observer based on inverting a simple rational model of action. However, in some domains, human action planning may be quite rational, and our inferences may be based on inverting near-ideal models of behavior. Two-dimensional navigation tasks, such as our maze-world stimuli, are a good candidate for a domain in which humans have evolved to act and interpret others' actions near-optimally (Barrett et al., 2005). Or maybe in some cases, as in intuitive physics (McCloskey et al., 1980), our mental models of the world are simpler than the world itself; if so, a probabilistic rational-actor model seems to be a reasonable first approximation for the mind to adopt in interpreting other people's behavior. Future work should further explore the correspondence between people's intuitive models of goal-directed action and the actual mechanisms by which humans produce action.

By applying our framework across different tasks and contexts, and exploring different models of people's goal representations, we found evidence for flexible use of different goal representations depending on what the task and context suggests. It is not yet clear whether the mental representations uncovered by our inverse planning models will generalize to explain human action understanding outside of these laboratory tasks. More work here is certainly needed. Our expectation, however, is that many areas of intuitive psychology will be usefully illuminated by ideal inference models that combine a probabilistic principle of rationality with increasingly powerful representations of intentional mental states.

Acknowledgments

This work was supported by AFOSR MURI contract FA9550-05-1-0321, the James S. McDonnell Foundation Causal Learning Collaborative Initiative and the National Science Foundation Graduate Fellowship (CLB). We thank Noah Goodman, Konrad Körding and Pat Shafto for helpful comments and discussions.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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., Tenenbaum, J. B., & Saxe, R. R. (2006). Bayesian models of human action understanding. In *Advances in Neural Information Processing Systems* (Vol. 18, pp. 99–106).

Barrett, H., Todd, P. M., Miller, G., & Blythe, P. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*(4), 313-331.

Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line drawings as three dimensional surfaces. *Artificial Intelligence*, *17*(1–3), 75–117.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* New York: Oxford University Press.

Bellman, R. (1957). *Dynamic programming.* Princeton, N.J.: Princeton University Press.

Bertsekas, D. P. (2001). *Dynamic programming and optimal control* (2nd ed., Vol. 2). Belmont, MA: Athena Scientific.

Bertsekas, D. P., & Tsitsiklis, J. (1996). *Neuro-dynamic programming.* Belmont, MA: Athena Scientific.

Blythe, P. W., Todd, P. M., & Miller, G. F. (1999). How motion reveals intention: categorizing social interactions. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 257–286). New York: Oxford University Press.

Bui, H., Venkatesh, S., & West, G. (2002). Policy recognition in the abstract hidden Markov model. *Journal of Artificial Intelligence Research*, *17*.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, *12*(3), 314–325.

Charniak, E., & Goldman, R. (1991). A probabilistic model of plan recognition. In *Proceedings of the tenth national conference on artificial intelligence* (pp. 160–165).

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.

Csibra, G., & Gergely, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. *Cognition*, *63*(2), 227–233.

Csibra, G., Gergely, G., Biró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*, 237-267.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453.

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

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of Royal Society of London, Series B: Biological Sciences*, *308*, 67–78.

Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, *13*, 227–303.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, *12*, 961–974.

Feinberg, E. A., & Shwartz, A. (Eds.). (2002). *Handbook of Markov decision processes: Methods and applications*. New York: John Wiley and Sons.

Filar, J., & Vrieze, K. (1997). *Competitive Markov decision processes*. New York: Springer.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, *44*(3), 283–296.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2 ed.). New York: Chapman & Hall.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 150–184). Oxford: Clarendon Press.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*, 755.

Gergely, G., Nádasdy, Z., Csibra, G., & Biró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*, 447–474.

Goldman, A. I. (2006). *Simulating minds*. Oxford University Press.

Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. In J. Bernardo, M. DeGroot, D. Lindley, & A. Smith (Eds.), *Bayesian statistics* (pp. 489–519). Valencia: Valencia University Press.

Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., et al. (2006). Intuitive theories of mind: A rational approach to false belief. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 1382–1390).

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

Guajardo, J. J., & Woodward, A. L. (2004). Is agency skin deep? surface attributes influence infants' sensitivity to goal-directed action. *Infancy*, *6*(3), 361–384.

Guestrin, C., Koller, D., Parr, R., & Venkataraman, S. (2003). Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research*, *19*, 399–468.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Asssociation for Computational Linguistics* (pp. 159–166).

Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences*, *101*(19), 7487–7491.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heider, F., & Simmel, M. A. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243–249.

Hernstein, R. J., & Prelec, D. (1991). Melioration: a theory of distributed choice. *The Journal of Economic Perspectives*, *5*(3), 137–156.

Jefferys, W., & Berger, J. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, *80*, 64–72.

Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze following in 12-month-olds. *Developmental Science*, *1*(2), 233–238.

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

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kautz, H., & Allen, J. (1986). Generalized plan recognition. In *Proceedings of the fifth national conference on artificial intelligence* (pp. 32–37).

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Körding, K. (1997). Decision theory: What "should" the nervous system do? *Science*, *318*(5850), 606–610.

Liao, L., Fox, D., & Kautz, H. (2004). Learning and inferring transportation routines. In *Proceedings of the nineteenth national conference on artificial intelligence* (pp. 348–353).

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 157–163). Morgan Kaufmann.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, *35*(4), 549–568.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science*, *210*(4474), 1139–1141.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*, 470–476.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850.

Ng, A. Y., & Russell, S. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 663–670).

Parr, R., & Russell, S. (1998). Reinforcement learning with hierarchies of machines. In *Advances in Neural Information Processing Systems 10* (pp. 1001–1007).

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, *98*, 137–155.

Puterman, M. L. (2005). *Markov decision processes: Discrete stochastic dynamic programming*. New York: Wiley Interscience.

Richards, W., Jepson, A., & Feldman, J. (1996). Priors, preferences, and categorical percepts. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (p. 93-122). Cambridge, England: Cambridge University Press.

Saxe, R. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences*, *9*(4).

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

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Schultz, J., Friston, K., Wolpert, D. M., & Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, *45*, 625–635.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Gaag, C. V. der, Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of Royal Society of London, Series B: Biological Sciences*, *358*(1430), 415–427.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*(2), 145–166.

Sodian, B., Schoeppner, B., & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development*, *27*, 31–41.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Tesauro, G. (1994). Td-gammon, a self-teaching backgammon program, achieves master-level play. *Neural Computation*, *6*(2), 215–219.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*, 943–951.

Tremoulet, P. D., & Feldman, J. (2006). The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. *Perception and Psychophysics*, *29*, 943–951.

Verma, D., & Rao, R. (2006). Goal-based imitation as probabilistic inference over graphical models. In *Advances in Neural Information Processing Systems* (Vol. 18, pp. 1393–1400).

Watkins, C. J. C. H. (1989). *Learning from delayed rewards.* Unpublished doctoral dissertation, Cambridge University, Cambridge, England.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

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

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*, 239–277.

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

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology*, *4*(12), 1–14.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*, 979–1008.