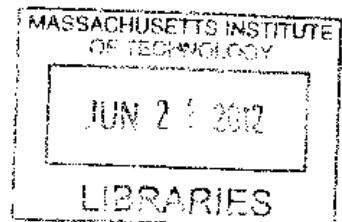


Bayesian Theory of Mind: Modeling Human Reasoning about Beliefs, Desires, Goals, and Social Relations

by

Chris L. Baker

B.S., University of Washington (2004)



Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

ARCHIVES

Doctor of Philosophy in Cognitive Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

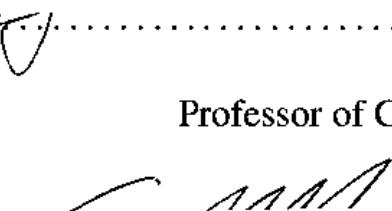
June 2012

© 2012 Massachusetts Institute of Technology. All rights reserved.

Author


Department of Brain and Cognitive Sciences

May 8, 2012

Certified by


Joshua B. Tenenbaum, PhD

Professor of Cognitive Science and Computation

Thesis Supervisor

Accepted by


Earl K. Miller, PhD

Picower Professor of Neuroscience and

Director, Department of Brain & Cognitive Sciences Graduate Program

Bayesian Theory of Mind: Modeling Human Reasoning about Beliefs, Desires, Goals, and Social Relations

by

Chris L. Baker

Submitted to the Department of Brain and Cognitive Sciences
on May 8, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis proposes a computational framework for understanding human *Theory of Mind* (ToM): our conception of others' mental states, how they relate to the world, and how they cause behavior. Humans use ToM to predict others' actions, given their mental states, but also to do the reverse: attribute mental states – beliefs, desires, intentions, knowledge, goals, preferences, emotions, and other thoughts – to explain others' behavior. The goal of this thesis is to provide a formal account of the knowledge and mechanisms that support these judgments.

The thesis will argue for three central claims about human ToM. First, ToM is constructed around probabilistic, causal models of how agents' beliefs, desires and goals interact with their situation and perspective (which can differ from our own) to produce behavior. Second, the core content of ToM can be formalized using context-specific models of approximately rational planning, such as Markov decision processes (MDPs), partially observable MDPs (POMDPs), and Markov games. ToM reasoning will be formalized as rational probabilistic inference over these models of intentional (inter)action, termed Bayesian Theory of Mind (BToM).

Third, hypotheses about the structure and content of ToM can be tested through a combination of computational modeling and behavioral experiments. An experimental paradigm for eliciting fine-grained ToM judgments will be proposed, based on comparing human inferences about the mental states and behavior of agents moving within simple two-dimensional scenarios with the inferences predicted by computational models.

Three sets of experiments will be presented, investigating models of human goal inference (Chapter 2), joint belief-desire inference (Chapter 3), and inference of interactively-defined goals, such as chasing and fleeing (Chapter 4). BToM, as well as a selection of prominent alternative proposals from the social perception literature will be evaluated by their quantitative fit to behavioral data. Across the present experiments, the high accuracy of BToM, and its performance relative to alternative models, will demonstrate the difficulty of capturing human social judgments, and the success of BToM in meeting this challenge.

Thesis Supervisor: Joshua B. Tenenbaum, PhD
Title: Professor of Cognitive Science and Computation

To my parents: Esther, Mike, Kathi, and Jess.

Acknowledgments

“You should not forget the feeling of gratitude. It is rare to meet one another and to practice what is rare to practice.”

– Zen Master Dōgen, *Regulations for the Auxiliary Cloud Hall*

I came to MIT to work with Josh Tenenbaum on the precise topic of this thesis: computational modeling of theory of mind. This was a somewhat risky proposition – a novel research direction for the lab, and for the field in general, but Josh convinced me that we would succeed. It is a testament to his penetrating vision, and to his skills as an advisor, how consistent this thesis is with some of our earliest discussions. Since then, I’ve realized that this was no fluke – Josh’s students have continued to embark on ambitious research expeditions, with impressive results every time. Josh really is that smart.

Rebecca Saxe played an integral role in this research program from day one. During my first semester, she and Josh organized the “Computational Mindreading Group” (other key members included Whitman Richards, Konrad Körding, and Kobi Gal). Her expertise on theory of mind research guided us toward the prior work, open questions, and theoretical approaches that continue to inspire my research today. As a collaborator, Rebecca’s input on all aspects of my research has been essential. Rebecca will always set a high standard for my future contributions to the field.

Leslie Kaelbling and Nancy Kanwisher made vital contributions as advisory committee members. Leslie played a particularly important role in encouraging and providing suggestions on the material in Chapter 3. Nancy ensured that my results were evaluated rigorously, while providing great enthusiasm and kindness throughout. Leslie and Nancy are inspiring scientists and outstanding role models for an aspiring academic.

Chapters 2 & 3 of this thesis are based on papers with Rebecca Saxe and Josh Tenenbaum, and Chapter 4 is based on a paper with Noah Goodman and Josh Tenenbaum. Noah Goodman has been a close collaborator on several other papers and projects, and always an excellent source of ideas and inspiration.

My additional fantastic coauthors include: Liz Baraff Bonawitz, Vikash Mansinghka, Alison Gopnik, Henry Wellman, Laura Schulz, Pat Shafto, Russell Warner, Tomer Ullman, Owen Macin-doe, Owain Evans, Julian Jara-Ettinger, Hilary Richardson, Kiley Hamlin, Jacob Feldman, and Peter Pantelis. Our collaborations have genuinely enhanced my thinking and benefitted every aspect of this thesis.

Many members and affiliates of the CoCoSci and Saxe labs provided feedback and friendship: Konrad Körding, Tom Griffiths, Charles Kemp, Tevye Krynski, Amy Perfors, Lauren Schmidt, Kobi Gal, Brian Milch, Virginia Savova, Dan Roy, Mike Frank, Ed Vul, David Wingate, Peter Battaglia, Liane Young, Marina Bedny, Emile Bruneau, Frank Jäkel, Ruslan Salakhutdinov, Yarden Katz, Steve Piantadosi, Hyowon Gweon, John McCoy, Tim O'Donnell, Brenden Lake, Jessica Hamrick, Cameron Freer, Andreas Stuhlmüller, Leon Bergen, Nick Dufour, Katherine Heller, Abe Friesen, Mina Cikara, Ben Deen, Jen White, Tao Gao, Josh Hartshorne, Jon Scholz, Max Siegel, Jon Malmaud and Eyal Dechter. I have also enjoyed working with a number of talented undergraduate researchers in the CoCoSci lab: Matthew Ng, Kevin Wang, Lindsay Johnson, Jessica Su and Sahar Hakim-Hashemi.

Professor Rajesh Rao, my undergraduate research advisor at the University of Washington was an exceptional mentor, providing my first exposure to computational neuroscience, machine learning, and social robotics. In Raj's lab, I worked closely with (then) PhD students Aaron Shon and David Grimes, who taught me an incalculable amount, and inspired me to follow in their footsteps by pursuing my PhD. Their work on Bayesian imitation learning has continued to have an important influence on my own research.

I owe special thanks to the hosts of several talks during 2011-2012 which helped sharpen the ideas in this thesis: Tamim Asfour, Chris Geib, Robert Goldman, Henry Kautz, Rineke Verbrugge, Jacob Feldman, and Camille Morvan. Thanks as well to the hosts and attendees of the MIT BCS CogLunch series and the MIT CSAIL ML-Tea series for tolerating several of my talks.

I am grateful to the MIT Leventhal Presidential Graduate Fellowship, the Department of Homeland Security Graduate Fellowship, and the National Science Foundation Graduate Fellowship for their generous funding. Additional funding was provided by AFOSR MURI contract FA9550-05-

1-0321, the James S. McDonnell Foundation Causal Learning Collaborative Initiative, and ARO MURI grant W911NF-08-1-0242. Through activities and meetings relating to these and several other grants, I have been fortunate to interact with many interesting colleagues who have had a lasting influence on my work: Tina Eliassi-Rad, David Roberts, Whitman Richards, Scott Atran, Brian Stankiewicz, Tamar Kushnir, Pedro Domingos, Tom Dietterich, Drew Bagnell, and Brian Ziebart.

The Department of Brain and Cognitive Sciences has been an amazing place to do my PhD. The friendships I've made at MIT – too many to list here – will last long after I've left. Special thanks to Denise Heintze, the BCS academic administrator, for administering my graduation.

Thank you to the Cambridge surf crew: Devin McCombie (Deuce), James Evans (3stroke), Dariusz Golda (Sarge), Alec Robertson (Buoy), Sebastian Sovero (Cbass), and Dan Sura (Prof. de las Olas) for teaching me to shred!

Thank you to Boardman house: Bren Galvez-Moretti, Matt Chua, Kyle Steinfeld, Dennis Ott, Slobodan Radoman, Ben Scott, Vivek Freitas, and Veronika and Dod Bernstein for being awesome.

This thesis is dedicated to my parents, but my entire family has always been exceptionally supportive of my education. Thank you to all my grandparents, siblings, aunts, uncles, and cousins who have helped me along the way.

Caroline Runyan has been my closest companion, supporter, friend, helper, and colleague for most my time at MIT. Her talent as a scientist is a constant inspiration. It has been truly special to work on our PhDs together and finish at the same time.

Contents

1	Introduction	15
1.1	Literature Survey	18
1.1.1	Comparative and Developmental Perspectives	19
1.1.2	Computational Perspectives	24
1.1.3	Cognitive Science Perspectives	28
1.2	Computational Framework	29
1.3	Experimental Paradigm	31
2	Action Understanding as Inverse Planning	33
2.1	Introduction	33
2.2	Computational Framework	40
2.2.1	Related Work	45
2.3	Experiment 1	46
2.3.1	Method	46
2.3.2	Results	50
2.3.3	Discussion	55
2.4	Experiment 2	57
2.4.1	Method	58
2.4.2	Results	59
2.4.3	Discussion	62
2.5	Experiment 3	63

2.5.1	Method	65
2.5.2	Results and Discussion	68
2.6	General Discussion	71
2.7	Conclusion	76
3	Bayesian Theory of Mind	79
3.1	Introduction	79
3.2	Computational Framework	81
3.2.1	Informal sketch	82
3.2.2	Formal modeling	85
3.2.3	Alternative models	86
3.3	Experiment	87
3.3.1	Results & Discussion	89
3.4	Conclusion & Future Work	91
4	Theory-based Social Goal Inference	93
4.1	Introduction	93
4.2	Computational framework	96
4.2.1	(Inverse) planning with social goals	97
4.2.2	Cue-based models	99
4.3	Experiment	100
4.3.1	Method	100
4.3.2	Results	103
4.4	Conclusion	103
5	Conclusion	107
A	Formal details of goal inference modeling	111
A.1	Markov decision problems	111
A.2	Inverse planning models	114

A.2.1	Model 1: single underlying goal	114
A.2.2	Model 2: changing goals	116
A.2.3	Model 3: complex goals	118
A.2.4	Heuristic alternative H	119
A.3	Experiment 1	120
A.3.1	Bootstrap cross-validated correlational analyses	120
A.3.2	Analysis of parameters	121
A.4	Experiment 2	122
A.4.1	Bootstrap cross-validated correlational analysis	122
A.4.2	Analysis of parameters	122
A.5	Experiment 3	123
A.5.1	Data processing and model fitting	123
A.5.2	Bootstrap cross-validated correlational analysis	123
A.5.3	Analysis of parameters	124
A.6	General discussion	125
A.6.1	Hierarchical Bayesian analysis: model selection across experiments . . .	125
References		127

Chapter 1

Introduction

“If we removed all knowledge of scientific psychology from our world, problems in interpersonal relations might easily be coped with and solved much as before. Man would still ‘know’ how to avoid doing something asked of him, and how to get someone to agree with him; he would still ‘know’ when someone was angry and when someone was pleased. He could even offer sensible explanations for the ‘whys’ of much of his behavior and feelings. In other words, the ordinary person has a great and profound understanding of himself and of other people which, though unformulated and only vaguely conceived, enables him to interact with others in more or less adaptive ways.”

— Fritz Heider, *The Psychology of Interpersonal Relations* (1958)

Making sense of everyday social situations requires us to think about the thoughts that cause others’ behavior—what beliefs, desires, goals, or feelings make people act the way they do? Though such questions confound scientific psychologists, for ordinary humans they are merely academic—as *intuitive psychologists*, we understand the answers with ease. Despite the notorious difficulty of knowing what someone else is thinking, adults (and even infants) are experts at it—we make complex social inferences rapidly, often unconsciously, and with impressive accuracy. How do we do it?

The classical response (e.g., Premack & Woodruff, 1978; Wimmer & Perner, 1983; Wellman, 1990; Gopnik & Meltzoff, 1997) is that humans use a *Theory of Mind* (ToM): an abstract, coherent body of knowledge about other people's (and our own) minds—how they are structured, how they relate to the world, and how they cause behavior. We know that others have beliefs about the world, which depend on their subjective interpretation of the information available; that they have desires for contingent circumstances; and that these mental states guide their actions in predictable ways, favoring the means most likely to achieve their desires, according to their beliefs.

Though fundamental, the insights provided by traditional accounts of ToM are restricted to qualitative phenomena—they do not explain quantitative variation in human reasoning as a function of actions, situations, or prior expectations of plausible mental states. Furthermore, generating empirical predictions from non-formal accounts involves introspecting on our own intuitive psychological knowledge – a circular reliance on the very intuitions theories of ToM must ultimately explain. How can we go beyond these limitations to develop quantitatively testable theories which express the structure and content of ToM in explicit, formal terms?

The purpose of this thesis is to define a computational framework for understanding the central concepts of human ToM, and to rigorously test its capacity to capture human judgments across a range of social situations. The thesis will argue for three primary claims—two theoretical, one methodological. First, ToM is constructed around probabilistic, generative, causal models of agents' mental states and behavior, which express how actions are caused by beliefs and desires¹, and how beliefs about the world are caused by perceptual access to the present situation. These models support prediction of agents' behavior, given their beliefs and desires, but also the reverse – causal inference of beliefs and desires, given observed behavior – through mechanisms of Bayesian reasoning.

Second, the content of ToM can be understood in terms of models of approximately rational belief-, desire-, and goal-dependent planning, formalized as maximization of expected gains and minimization of expected losses. This frames ToM as a kind of “intuitive economics” – a folk

¹These are the so-called “propositional attitudes”, central to classical discussions of ToM and “folk-psychology” (e.g., Wimmer & Perner, 1983; Dennett, 1987; Churchland, 1981). This thesis focuses on beliefs and desires; in general, ToM may encompass a broader class of mental states, e.g., including emotions (although the extent to which the propositional content of these may be reducible to beliefs and desires is debatable (e.g., Searle, 1983)).

analog of expected utility theory used for understanding everyday behavior. Importantly, this does not assume that humans necessarily plan or act rationally in all circumstances. Rather, we tacitly expect others to behave rationally (or approximately so) in particular situations, and perform “explanation by rationalization” by attributing the mental states which make their behavior appear most rational within the present context.

The Bayesian Theory of Mind (BToM) framework combines these principles governing the structure and content of ToM, inverting models of planning and reasoning using Bayesian inference to yield quantitative predictions of human judgments. The third claim of this thesis is that different hypotheses about the structure and content of human ToM can be tested by combining computational modeling with behavioral experiments. This thesis proposes a paradigm for large-scale measurement and modeling of quantitative psychophysical data, based on comparing human inferences about the mental states of agents moving within simple two-dimensional scenarios with the inferences predicted by computational models. Quantitative comparison of human judgments with computational models defined over different knowledge representations – different priors over kinds of mental states and relations among them – provides a basis for exposing the structured knowledge comprising human ToM.

BToM will be applied in three experimental scenarios. In each scenario, BToM uses an account of rational agency (adapted from Economics and Computer Science) to capture people’s context-sensitive knowledge and reasoning. First, human goal attribution will be formalized as Bayesian inference over Markov decision process (MDP) models of goal-directed action. Second, joint belief-desire attribution will be formalized as Bayesian inference over partially observable MDP (POMDP) models of belief- and desire-dependent planning. Third, attribution of social goals to interacting agents, such as chasing and fleeing, will be modeled as Bayesian inference over Markov game models of multiagent interaction. The performance of BToM will be compared to a class of discriminative behavioral classifiers based on low-level motion cues, demonstrating the advantage of the BToM account.

This chapter will begin by sketching the history of research on ToM and social reasoning, focusing on attempts at a formal characterization. Next, the proposed computational framework

will be sketched at a high level, in order to convey the general modeling approach. Lastly, an experimental paradigm for studying ToM will be described.

1.1 Literature Survey

The study of the human conception of other minds is as old as psychology itself; consider William James' (1890) famous declaration in *The Principles of Psychology*: “Everyone knows what attention is”—or Gilbert Ryle: “We possess already a wealth of information about minds, information which is neither derived from, nor upset by, the arguments of philosophers” (1949). Indeed, everyone knows what minds are like, through introspective access to our own mental states, and reasoning about others’ thoughts and feelings during everyday social interaction. On this view, early philosophical, psychological, and economic theories can be regarded as codifying our common knowledge and intuition about the mind in earnest detail (e.g., Descartes, 1641; Hume, 1739; Mill, 1874; James, 1890; Pareto, 1906).

Fritz Heider termed this knowledge “common-sense or naive psychology”, defined as our lay understanding of other people, their thoughts, their actions, and how these depend on their perceived situation (1958)². Similar to object perception, social perception often involves sophisticated inferences which go far beyond the data observed, explaining a person’s behavior in terms of causal factors, such as their beliefs, desires, intentions, emotions, dispositions, perceptions, experiences, expectations, and subjective conception of the environment. Heider proposed various principles for performing these attributions, the most well-known being the schema of *equifinality*: “the invariance of the end and variability of the means” as a distinguishing feature of intentional causality (1958). Heider’s explicit framing of social reasoning as a problem of causal understanding was a seminal contribution to the field of social psychology. However, critical challenges arise from this perspective. First, how can we build on Heider’s insights about the structure, content and function of social knowledge? Second, how can we go beyond limited principles such as equifinality to give a more general account of social reasoning and inference?

²Heider argued that although naive psychology may be incorrect (cf. Churchland, 1981), it is nevertheless a fundamental part of our human cognitive endowment that determines how we reason about and interact with others.

A fresh approach to these problems emerged with a classic paper by Premack and Woodruff (1978) on Chimpanzees' ability to reason about others' behavior and mental states. This paper was the first to characterize the core representations of commonsense psychology as a "Theory of Mind" (ToM): structured knowledge about mental states (such as "*purpose* or *intention*, as well as *knowledge*, *belief*, *thinking*, *doubt*, *guessing*, *pretending*, *liking*, and so forth" (Premack & Woodruff, 1978)) that are inferred to explain and predict behavior. The ensuing commentary by several philosophers (Dennett, 1978; Bennett, 1978; Harman, 1978), who had explored related ideas around this time (Churchland, 1970; Dennett, 1971; D. Lewis, 1972; Dennett, 1987, chap. 3) under the term "folk psychology", inspired a stronger criterion for the possession of ToM, less sensitive to non-mentalistic interpretations, known as the "Sally-Anne" false-belief task (Wimmer & Perner, 1983). In this task, children are told a story in which Sally leaves her marble in a box, then goes outside; while she is away, Anne moves Sally's marble to a basket. Children are then asked to predict where Sally will look for her marble when she returns. Wimmer and Perner (1983) found that 3 year-old children overwhelmingly gave the incorrect, reality-centric response (basket), only responding correctly based on Sally's putative belief (box) around age 4 or 5. Crucially, answering correctly requires representing Sally's belief, considered a sufficient condition for possessing a ToM; the correct answer cannot be determined on the basis of the participant's own knowledge about the location of the marble. A huge body of subsequent research has studied false-belief understanding in the Sally-Anne and other false-belief tasks (e.g., Flavell, Flavell, and Green (1983); Perner and Wimmer (1985); Hogrefe, Wimmer, and Perner (1986); Perner, Leekan, and Wimmer (1987); Gopnik and Astington (1988); Perner, Frith, Leslie, and Leekman (1989)). By all of these measures, children's ability to represent false beliefs develops around the fourth year of life (Wellman, Cross, & Watson, 2001); autistic children's much later (Baron-Cohen, Leslie, & Frith, 1985); and great apes' perhaps never (Call & Tomasello, 1999).

1.1.1 Comparative and Developmental Perspectives

As suggested by the preceding discussion, a powerful way to elucidate the knowledge and reasoning mechanisms underlying social intelligence is by studying their development and evolution.

Social development is both a constructive and destructive process, building on top of, modifying, and replacing earlier representations and mechanisms (Piaget, 1936, 1937; Siegler, 1996; Carey, 1985; Wellman & Gelman, 1992; Gopnik & Meltzoff, 1997). As Piaget argued, by careful study of continuities and discontinuities throughout development, the ontogenesis and structure of human knowledge can be ascertained³. In the social domain, Piaget argued that young children are fundamentally “egocentric”, hallmarked by their inability to take another person’s perspective. For example, when four year-old children performing the “three mountains task” selected from a set of pictures to describe how another person would see a 3D mountain display from various different viewpoints, they usually chose the picture taken from their own viewpoint (Piaget & Inhelder, 1948). Six year-olds often chose pictures of different viewpoints, but rarely the correct one. Only at ages 7-8 could children reliably succeed at the task. Piaget interpreted these results in terms of the transition from the “pre-operational” stage to the “concrete-operational” stage (Piaget, 1937).

Since Piaget, developmental psychologists have discovered that infants are innately predisposed toward social interaction, learning and understanding. Using clever new experimental methods matched to younger children’s comprehension and response capabilities, sophisticated knowledge and capacities have been shown to develop dramatically earlier than Piaget claimed across a range of cognitive domains (e.g., Kellman and Spelke (1983); Baillargeon, Spelke, and Wasserman (1985); Luo and Johnson (2009)). For social cognition, evidence is rapidly accumulating for the emergence of fundamental mentalistic concepts and reasoning within the first year of life. Currently the specific ages at which these essential capacities develop, and the precise order in which they do so, are moving targets (downward). We may eventually know the true developmental picture, but the outlines are already becoming increasingly clear: an amazingly coherent network of social concepts is within present in humans within the first year of life, and provides a foundation for social cognition throughout the lifespan.

A primary goal of this thesis is to provide an account of the nonverbal social knowledge and reasoning that emerges within infants’ first year of life. This section reviews the development of

³Lev Vygotsky, another pioneering developmental psychologist, argued for a somewhat different perspective, essentially that children’s knowledge develops through mechanisms of *internalization* and *appropriation* of social-interactive behavior and skills (Vygotsky, 1934).

three central ToM capacities: reasoning about motivation, such as desires, intentions and goals; reasoning about epistemic mental states, such as belief and knowledge; and reasoning about interactively defined mental states – social goals like “to help” or “to hinder”. First, by 6 months old, infants already use simple representations of agents, actions and goals to process observations of behavior. A classic experiment by Woodward (1998) habituated 5- and 6-month-old infants to an arm reaching from behind a curtain to grasp one of two objects. The objects were always in the same locations on each trial, and the arm always grasped the same object. In a testing phase, the objects’ locations were swapped, and infants’ looking time was measured in response to either the arm to grasping for the same object in its new location, or the other object which was now in the location of the previously grasped object. Infants expected the arm to grasp the same object, looking longer when the other object was grasped, consistent with the attribution of an object-directed goal toward a specific object that was invariant to its location. In several control conditions, when infants were habituated to the same movements being made by various non-human appendages (e.g., an arm-shaped pole), they looked the same amount of time in both testing conditions, showing that they genuinely attributed a goal to the human agent, and were not simply forming an association between movements and objects⁴. Infants’ own grasping ability typically develops only slightly before 5 months. In an ingenious study, by first training younger infants to grasp objects using special “sticky mittens”, Sommerville, Woodward, and Needham (2005) showed that infants could then use this knowledge to infer the goals of others’ grasping actions at just 3 months old.

Before their first birthday, infants’ perception of intentional action increases swiftly in representational and inferential power, allowing them to reconstruct mental states from incomplete observations. For example, Hamlin, Hallinan, and Woodward (2008) showed that 7-month-old infants imitate both successful and unsuccessful object-directed grasps, suggesting that they can infer the goals of actions even without observing the intended outcome directly. In a control experiment, infants did not reliably imitate the selection when instead of a grasp, contact with the object was made with the back of the hand, indicating their sensitivity to the appropriateness of the action

⁴Under certain circumstances, 3-month-olds (Luo, 2011), 5-month-olds (Luo & Baillargeon, 2005) and 6.5-month-olds (Csibra, 2008) have been shown to attribute goals to non-human agents; this tendency increases with age (e.g., Gergely and Csibra (2003); Heider and Simmel (1944)).

being performed. In a series of studies, Gergely, Csibra and colleagues showed that 9- (Csibra, Gergely, Biró, Koós, & Brockbank, 1999) and 12-month-old (Gergely, Nádasdy, Csibra, & Biró, 1995; Csibra, Biró, Koós, & Gergely, 2003) infants apply an intuitive schema relating actions, goals, and the environment to interpret non-human agents' motion in 2-dimensional animated displays. 12-month-olds were shown to use this schema productively: given observations of any two of an agent's actions, goal, or environment, infants can infer the third (Csibra et al., 2003; Wagner & Carey, 2005).

Second, 6-month-old infants are sensitive to the way that knowledge depends on perceptual access to the environment, and integrate this information with goal or preference attributions to form expectations about behavior. For example, Luo and Johnson (2009) created a variant of the task of Woodward (1998) (described above) that included a condition in which one of the objects could be occluded from the human actor during the habituation phase. In the testing phase, the objects' locations were swapped, and now both objects were visible to the actor. Infants' looking times suggested that they only attributed a goal to the actor when both objects had been visible during the habituation phase, looking longer when the other object was chosen in this condition (replicating the corresponding condition of Woodward (1998)). In the condition where the other object was occluded during the habituation phase, infants did not infer a goal (more accurately a preference) for either object, consistent with the understanding that the previous choice did not reflect a preference because the actor didn't know the occluded object was there. Recent evidence has indicated that false-belief understanding is present as early as 10 months of age (Luo, 2011); it is almost certainly in place by the beginning of the first year (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; Luo & Baillargeon, 2007).

Third, infants attribute "social goals" to interacting agents by 6 months old, for instance, whether one agent is helping or hindering another. In several experiments, Hamlin and colleagues have shown that 5- (Hamlin & Wynn, 2011), 6-, and 10-month-old (Hamlin, Wynn, & Bloom, 2007) infants form preferences based on their evaluation of an agent's prosociality, neutrality, or antisociality toward other agents, choosing helpers over neutral agents and hinderers, and neutral agents over hinderers. This pattern was also partially evident in a follow-up study with 3-month-

olds (Hamlin, Wynn, & Bloom, 2010), in which infants looked longer at helpers than hinderers (consistent with prior results showing that infants look longer at agents for whom they have a social preference (Kuhlmeier, Wynn, & Bloom, 2003)). However, a control experiment showed that while infants looked longer at neutral agents than hinderers, they looked the same amount at helpers versus neutral agents. These results suggest a “negativity bias” in 3-month-olds’ evaluations or preferences—although they evaluated the hinderer negatively, and were averse to looking at antisocial agents, they either did not evaluate helpers positively, or did not form a consistent preference between helpers and neutral agents at such a young age (Hamlin et al., 2010).

Based on these phenomena, Spelke and Kinzler (2007) and Carey (2011) propose that reasoning about agents and actions is the function of a “core system” in human cognition. An influential view among developmental psychologists and philosophers is that ToM consists of an intuitive *causal* schema (Fig. 1-1) in which beliefs, desires, and intentions generate actions via abstract causal laws (Gopnik & Meltzoff, 1997; Wellman & Bartsch, 1988; Wellman & Gelman, 1992; Luo & Baillargeon, 2010). On this view, social attribution involves causal inference, working backward from observed effects (actions) to underlying causes (mental states). Young children’s social inferences reflect the development of this schema, which is initially defined over non-representational content (actions, goals, environment), and later extended to representational content such as beliefs, desires and intentions (Wellman & Bartsch, 1988; Perner, 1991; Leslie, 1995; Bartsch & Wellman, 1995; Csibra et al., 1999; Scott & Baillargeon, 2009). Dennett (1987) argues that the content of ToM can be expressed in terms of the *intentional stance*: the assumption that agents will have sensible beliefs about the world and their situation, and will act rationally to achieve their desires, given their beliefs. In Fig. 1-1, the intentional stance is expressed in terms of the *Principle of rational belief*, relating the Situation, Observation and Belief nodes, and the *Principle of rational action*, relating Belief, Desire and Action. The *teleological stance* proposed by Gergely et al. (1995) is a non-representational version of the intentional stance, which posits shared causal principles between young infants’ non-representational reasoning about agents and actions, and older children and adults’ representational ToM.

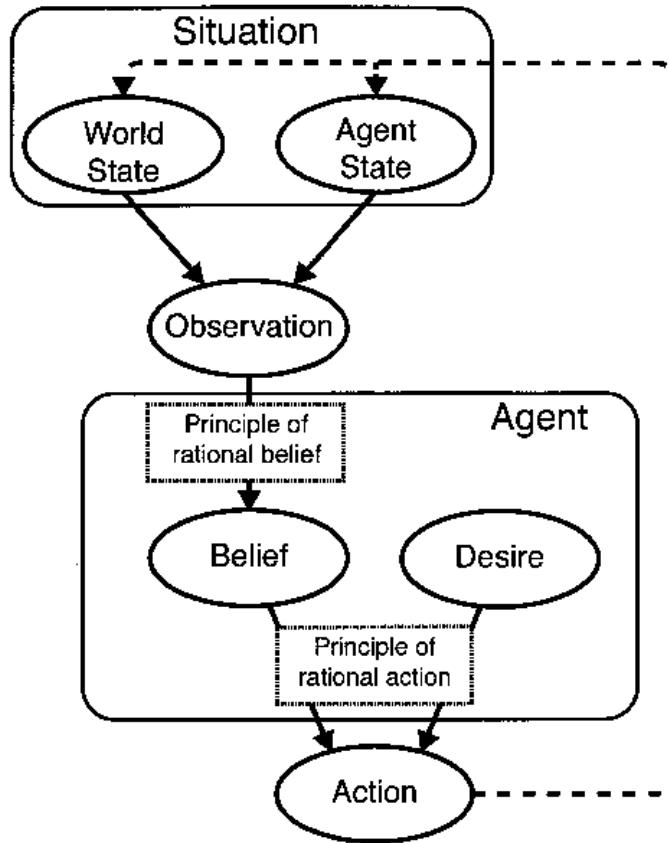


Figure 1-1: Causal schema for social reasoning, illustrating an *observer's* representation of an *agent's* mind, situation and behavior. Throughout the proposal, the terms “agent” and “observer” will be used to refer to these respective roles. Versions of this picture have been suggested by many authors (e.g., Wellman and Bartsch (1988); Baron-Cohen (1995); Brafman and Tennenholtz (1997); Csibra et al. (2003); Nichols and Stich (2003); A. I. Goldman (2006); Goodman et al. (2006)). Although though these accounts assume similar structural principles, the content of the knowledge they posit can be very different, e.g., “intuitive theory” (Wellman & Bartsch, 1988) versus “mental simulation” (A. I. Goldman, 2006).

1.1.2 Computational Perspectives

Computational modeling of commonsense psychology has a long history in the behavioral sciences. Key dimensions of variation among different approaches have been (1) whether they represent a concept of intentional agency, (2) the degree to which they support flexible, inductive inference and learning of unobserved mental states, and (3) the richness of the intuitive psychological theory they express. Table 1.1 summarizes these features for the selection of relevant literatures

and approaches discussed in this section.

Heider's followers in social psychology formulated statistical models of causal attribution: determining which of multiple potential causes (usually divided into "personal" and "impersonal" categories) produced an observed behavior. Harold Kelley proposed that causation can be attributed to the person or situation with which an effect *covaries* over time, and that naive observers apply the covariation principle using qualitative versions of the "analysis of variance" (ANOVA) procedure used in scientific data analysis (Kelley, 1972a, 1973). When an effect is observed only once, covariation cannot be assessed. In this situation, Kelley proposed that attributions can be performed using *causal schemata* (Kelley, 1972b, 1973). A causal schema determines which configurations of causes are *a priori* assumed capable of producing an effect. The *discounting* principle decreases the perceived responsibility of individual causes when other sufficient causes are present, while the *augmentation* principle increases the perceived responsibility if inhibitory causes are present. These principles allow observers to infer the presence or absence of unobserved causes, for example the skill level of a person who succeeds at a difficult task.

Kelley's formulation of social attribution as an inductive, statistical inference was an important advance toward understanding possible psychological mechanisms of social reasoning. Early attempts were made toward a general Bayesian formulation of these principles (Ajzen & Fishbein, 1975); these ideas have been recently revisited by Griffiths and Tenenbaum (2009); Kemp, Goodman, and Tenenbaum (2007), who provide powerful methods for learning causal theories and causal schemata. However, in focusing on elementary relations between causes and effects, attribution theory neglects the core representation of intentional agency at the heart of theory of mind (Malle, 2004).

Classical philosophical discussions of commonsense psychological knowledge were more mindful of the rich semantics of mental representation, typically expressed in terms of "propositional attitudes" (Russell, 1956; Quine, 1956). Major advances in the theory of modal logics for expressing propositional attitudes were made by C. I. Lewis and Langford (1932); Carnap (1942); Kripke (1959), and applied to representing knowledge and belief by Hintikka (1962). Logicians and artificial intelligence researchers following this tradition (Cohen & Levesque, 1990; Fagin,

Halpern, Moses, & Vardi, 1995; Milch & Koller, 2000; van Benthem, 2001; van Ditmarsch, van der Hoek, & Kooi, 2006) have developed formal theories of rational agency using modal logics of knowledge, belief, desire and intention, which can be used to express and reason about complex mental states and actions.

Philosophical accounts of intentionality have been influential in research on pragmatics and discourse understanding in Linguistics (e.g. Cohen, Morgan, and Pollack (1990)). Inspired by speech act theory (Austin, 1962; Searle, 1969) and Gricean pragmatics (Grice, 1989), these approaches assume that speakers have a representation of listeners' knowledge, beliefs, desires and intentions (expressed in modal logic), and use these representations to plan utterances to influence interlocutors' mental states (Cohen & Perrault, 1979). The listener's interpretation of an utterance's meaning involves heuristic, rule-based reasoning about the speaker's beliefs, desires, intentions and planning process (Schank & Abelson, 1977; Perrault & Allen, 1980; Allen & Perrault, 1980; Grosz & Sidner, 1986; Litman & Allen, 1987)⁵. Connections of these ideas to commonsense psychology were explored under the term "plan recognition" (Schank & Abelson, 1977; Schmidt, Sridharan, & Goodson, 1978; Whiten, 1991), for which a general algorithmic approach was developed by Kautz and Allen (1986). In addition, the well-known "Belief-Desire-Intention" architecture for resource-bounded planning was an attempt to capture commonsense intuitions about intelligent agency (and to justify their rationality) (Bratman, 1987; Bratman, Israel, & Pollack, 1988); this framework was formalized by Cohen and Levesque (1990). However, a major drawback of all of these models was their inability to represent and manage uncertainty, which limited their capacity to capture the flexible inductive learning and reasoning found in human plan recognition and communication⁶.

A new generation of models formulated plan recognition as a problem of probabilistic inference over context-dependent, generative models of actions (Charniak & Goldman, 1993; Albrecht, Zukerman, & Nicholson, 1998; R. P. Goldman, Geib, & Miller, 1999; Pynadath & Wellman, 2000; Bui, Venkatesh, & West, 2002; Geib & Goldman, 2009). Unlike the previous generation of heuris-

⁵ Analyses based on sophisticated notions of *common knowledge* (D. Lewis, 1969; Fagin et al., 1995) and *joint intention* (Searle, 1990; Grosz & Sidner, 1990), are outside of the scope of this survey, but will be revisited in the Discussion chapter of the dissertation.

⁶Recent research has begun to formulate probabilistic versions of these ideas from pragmatics and discourse understanding (e.g. Frank, Goodman, Lai, and Tenenbaum (2009)).

tic approaches to plan recognition, these models did not assume that agents' actions were generated by planning. Instead, they either assumed that probability distributions over intention- (or goal-) dependent actions were specified *a priori* (Charniak & Goldman, 1993; R. P. Goldman et al., 1999; Pynadath & Wellman, 2000; Bui et al., 2002; Geib & Goldman, 2009), or that they were learned from a corpus of previously observed behavior (Albrecht et al., 1998). Given observed actions, the plan, goal, or intention most likely to have generated them was inferred using Bayes' rule. Recent research has applied probabilistic discriminative models to activity recognition (Liao, Fox, & Kautz, 2004; Vail, Veloso, & Lafferty, 2007), learning the structure and parameters of activities directly from behavioral data (Liao et al., 2004). However, neither generative nor discriminative probabilistic plan recognition models contain a representation of intent- or goal-directed planning, limiting their effectiveness in explaining purposeful, adaptive behavior, particularly in novel contexts.

A natural approach for integrating intentionality and probability is provided by utility theoretic definitions of rationality. Within the field of Economics, the seminal "Bayesian games" framework (Harsanyi, 1967, 1968a, 1968b) and research in behavioral economics and behavioral game theory (e.g., Stahl and Wilson (1995); Ho, Camerer, and Weigelt (1998); Camerer (2003); Camerer, Ho, and Chong (2004)) have modeled the degree to which people represent and reason about other players' types, preferences, and strategic reasoning processes. Research in the fields of multiagent systems and human-computer interaction has also explored these issues (Gmytrasiewicz & Durfee, 1995; Milch & Koller, 2000; Gal & Pfeffer, 2004; Gmytrasiewicz & Doshi, 2005; Pynadath & Marsella, 2005; Gal & Pfeffer, 2008; Doshi, Qu, Goodie, & Young, 2010).

Within the field of Machine Learning, the problem of inferring an agent's utility function to explain its behavior was formulated in terms of "inverse reinforcement learning" (IRL) by Ng and Russell (2000) (see also Chajewska and Koller (2000)). Reinforcement learning describes the problem facing an agent figuring out how to act to maximize reward in an unknown environment (Sutton & Barto, 1998). Inverse reinforcement learning is the opposite problem: given data on how an agent behaves, determine the unknown reward function it is maximizing. In recent years, IRL (a.k.a. "inverse optimal control") has seen increasing interest in machine learning (Chajewska,

Koller, & Ormoneit, 2001; Abbeel & Ng, 2004; Ramachandran & Amir, 2007; Neu & Szepesvári, 2007; Dvijotham & Todorov, 2010; Ziebart, Bagnell, & Dey, 2010; Waugh, Ziebart, & Bagnell, 2011; Choi & Kim, 2011); its burgeoning influence in Cognitive Science is described in the next section.

Perspective	Selected Reference	Period	Intentional agency	Probabilistic inference	Mentalistic representation
Common-sense Psychology	Heider (1958)	1940s-1950s	Yes	No	Yes
Causal Attribution	Kelley (1973)	1960s-1970s	No	Yes	Yes
Epistemic logic	Hintikka (1962)	1960s-2000s	Yes	No	Yes
Bayesian games	Harsanyi (1967)	1960s-	Yes	Yes	No
Script recognition	Schank and Abelson (1977)	1970s	Yes	No	Yes
Classical plan recognition	Kautz and Allen (1986)	1980s	Yes	No	No
Probabilistic plan recognition	Charniak and Goldman (1993)	1990s-2000s	No	Yes	No
Inverse reinforcement learning	Ng and Russell (2000)	2000s	Yes	Yes	No
Behavioral game theory	Camerer et al. (2004)	2000s	Yes	Yes	No
Bayesian theory of mind	Baker (2012)	2000s-2010s	Yes	Yes	Yes

Table 1.1: Computational perspectives on social reasoning relating to commonsense psychology and theory of mind.

1.1.3 Cognitive Science Perspectives

An array of computational approaches to action understanding have been explored in the field of Cognitive Science. “Rule-based” analyses of false-belief reasoning were proposed by Schmidt and Marsella (1991); T. R. Schultz (1991); Wahl and Spada (2000); Bello and Cassimatis (2006a, 2006b); Bello, Bignoli, and Cassimatis (2007); Bello (2011), and connectionist models were proposed by Berthiaume, Onishi, and Shultz (2008); Van Overwalle (2010). Models of discriminative

action categorization based on the analysis of low-level motion cues were proposed by Blythe, Todd, and Miller (1999); Barrett, Todd, Miller, and Blythe (2005); Zacks (2004). These approaches will be contrasted with the proposed framework throughout the dissertation. Most closely related to the present thesis, models of motor control have been applied to capture “simulation theoretic” approaches to understanding and imitation of physical movement (Wolpert, Doya, & Kawato, 2003; Oztop, Wolpert, & Kawato, 2005; Rao, Shon, & Meltzoff, 2007).

In recent years, a rapidly growing body of research has suggested that human judgments about intentional agents’ mental states (goals, intentions, preferences, desires, knowledge, beliefs) can be modeled as probabilistic inverse planning, or inverse decision-making: Bayesian inferences over predictive models of rational, intentional behavior (Baker, Tenenbaum, & Saxe, 2006; Verma & Rao, 2006; Goodman et al., 2006; Baker, Tenenbaum, & Saxe, 2007; Baker, Goodman, & Tenenbaum, 2008; Yoshida, Dolan, & Friston, 2008; Baker, Saxe, & Tenenbaum, 2009; Goodman, Baker, & Tenenbaum, 2009; Lucas, Griffiths, Xu, & Fawcett, 2009; Ullman et al., 2009; Bergen, Evans, & Tenenbaum, 2010; Yoshida, Seymour, Friston, & Dolan, 2010; Tauber & Steyvers, 2011; Jern & Kemp, 2011; Jern, Lucas, & Kemp, 2011). In these models, “planning” captures an intuitive, probabilistic version of the classical rational agent who updates their beliefs to conform with their observations and chooses sequences of actions expected to achieve their desires. This is the approach that the present thesis will adopt; its properties are described below.

1.2 Computational Framework

This thesis presents Bayesian theory of mind (BTOM): a theory-based Bayesian framework (Tenenbaum, Griffiths, & Kemp, 2006) for social reasoning. At the highest level of abstraction, the framework defines context-dependent structural and functional principles, which generate causal models appropriate to the given social context. Three schemata for social inference are shown in Fig. 1-2; these schema will correspond to thesis chapters on goal inference (Chapter 2), joint belief-desire inference (Chapter 3), and social goal inference (Chapter 4), respectively. Each schema defines a fixed causal structure, but the *content* of the nodes in the network, and the probabilistic relations between the nodes vary depending on the context-specific functional form. These functional

forms correspond to models of intentional agency, indicated by the boxes marked “MDP Planning”, “POMDP Planning” and “Bayesian belief updating” in Fig. 1-2.

Within their specified context, each of these models can be used to probabilistically generate (sample) sequences of mental states and behavior. Given knowledge of agents’ mental states and situation, this allows those agents’ actions to be predicted. These models also support Bayesian inference of agents’ mental states, given observations of their actions in some situation. Action prediction and mental state inference are both fundamental capacities of ToM, and both will be tested in this thesis.

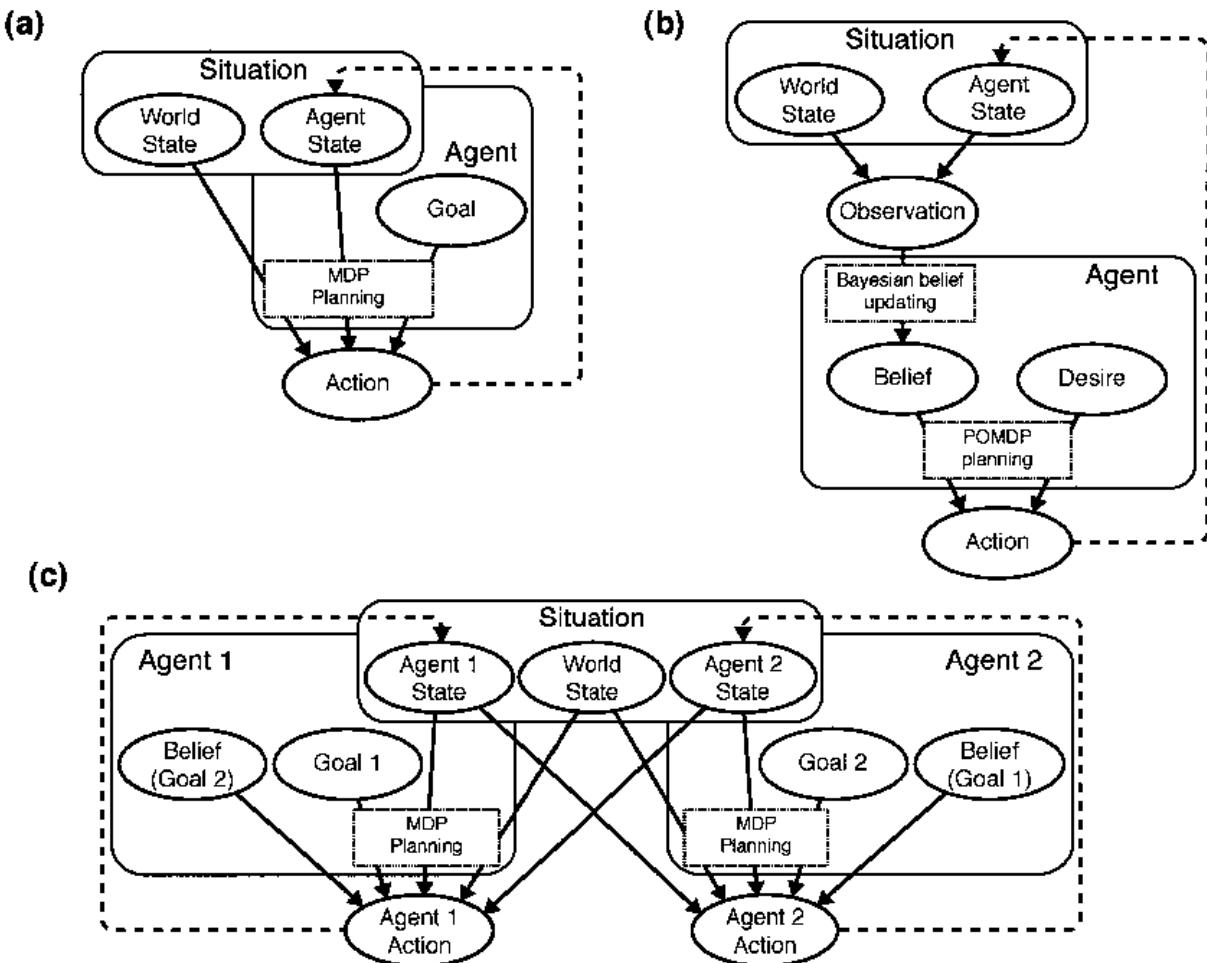


Figure 1-2: Schemata for social inference in theory of mind. For simplicity, all schema shown assume a static World State; agents’ actions are assumed to only change their own state (e.g., location in the environment). (a) Goal inference schema. (b) Belief and desire inference schema. (c) Social goal inference schema.

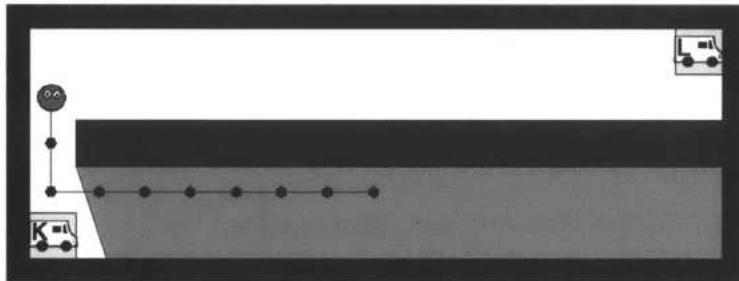
1.3 Experimental Paradigm

All experiments described in this thesis collected participants' inferences about agents' mental states in response to animations of agents moving in 2D grid-worlds and (example stimuli are shown in Fig. 1-3). Three sets of experiments will be described. The first set of experiments measured human inferences about agents' goals from observing their movements in simple spatial contexts (Fig. 1-3(a)). The second set of experiments elicited joint inferences of beliefs and desires in environments where agents could be uncertain about the state of the world, and gain information through line-of-sight visual access (Fig. 1-3(b)). The third set of experiments presented animations of multiple agents interacting, and collected inferences about agents' intentions toward one another (Fig. 1-3(c)). For each experiment, the raw stimulus time series was given as input to the model, and model predictions were compared with human judgments. This measured how closely human social inferences approach the ideal limit prescribed by a Bayesian observer. In addition, alternative models intended to capture competing theoretical accounts were formulated and tested as well. This determined whether the mentalistic reasoning captured by BToM is necessary to explain people's social inferences, or whether they can be accounted for by alternative theories.

(a)



(b)



(c)

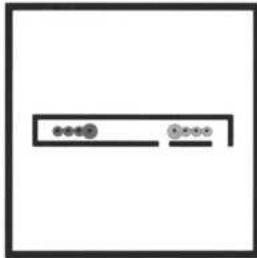


Figure 1-3: Stimuli from three different experimental scenarios. Experimental stimuli vary in complexity and fictional setting, but all include one or more agents moving in a 2D world among obstacles and sometimes goal objects. **(a)** Goal inference stimulus. Subjects judged whether an animated agent's goal was "A", "B", or "C" based on its trajectory. **(b)** Joint belief-desire inference stimulus. Subjects inferred agents' degree of desire for different trucks, and degrees of belief about different trucks' locations based on their movements. **(c)** Social goal inference stimulus. Subjects judged whether each agent was "chasing" or "fleeing" the other.

Chapter 2

Action Understanding as Inverse Planning

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.

2.1 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, Dayan, & Montague, 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, Niv, & Dayan, 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. 2-1(a) illustrates a typical example: a folk theory that specifies intentional agents’ beliefs and desires as the causes of their behavior (cf. 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, Jepson, & Feldman, 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 of behavior are qualitatively consistent with the inverse planning view (Meltzoff, 1995; Gergely et al., 1995; Meltzoff, 1988; Gergely, Bekkering, & Király, 2002; Csibra et al., 2003; Sodian, Schoeppner, & Metz, 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,

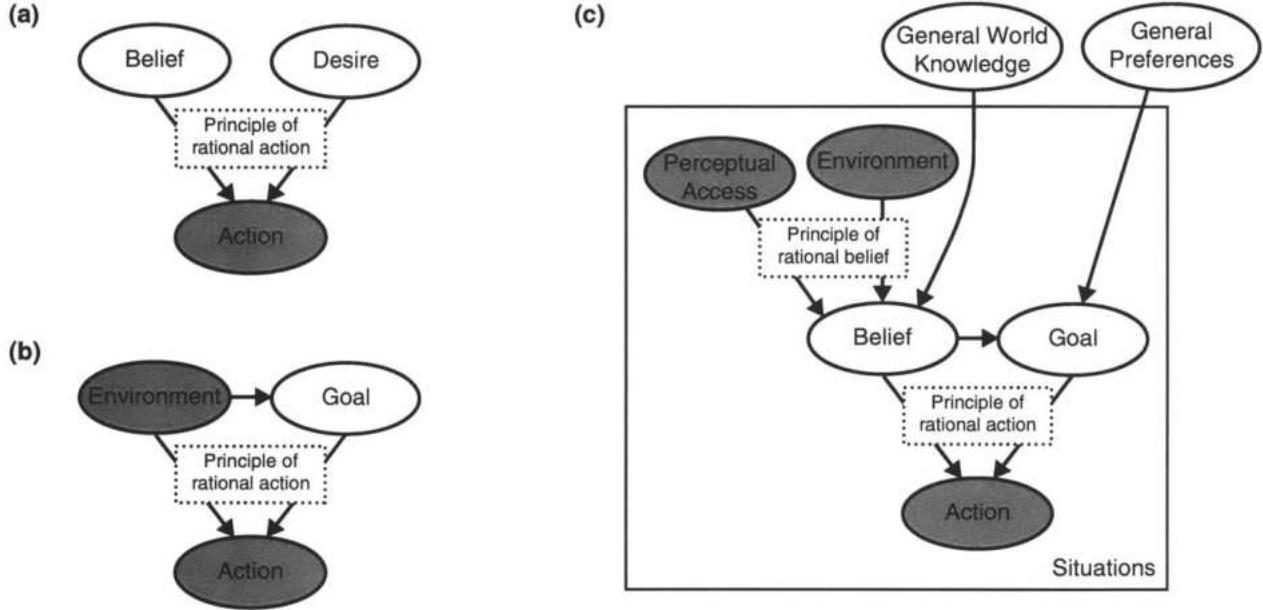


Figure 2-1: Modeling intuitive theories of intentional action. Diagrams use causal graph notation. Shaded nodes represent observed variables; unshaded nodes represent latent variables whose values must be inferred. Dotted boxes indicate the causal relation between variables. In this example, the observations 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**.

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. 2-1(a), shown in Fig. 2-1(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. 2-1(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. 2-1(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. 2-1(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. 2-1(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. 2-1(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, Durgin, & Kaufman, 1995; Scholl & Tremoulet, 2000; Tremoulet & Feldman, 2000; Zacks, 2004; R. T. Schultz et al., 2003; J. Schultz, Friston, Wolpert, & Frith, 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, Happe, Frith, & Frith, 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.

2.2 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 ac-

tion to the goal from every state. Our models assume that agents choose the optimal action only probabilistically. This yields a probability distribution over **Actions**, given a **Goal** and the **Environment**, denoted $P(\text{Actions}|\text{Goal}, \text{Environment})$, which provides the functional form of the probabilistic planning relation in Fig. 2-1(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 **Actions** and the **Environment**, using Bayes’ rule:

$$P(\text{Goal}|\text{Actions}, \text{Environment}) \propto P(\text{Actions}|\text{Goal}, \text{Environment})P(\text{Goal}|\text{Environment}). \quad (2.1)$$

In this equation, $P(\text{Actions}|\text{Goal}, \text{Environment})$ is the likelihood of the **Goal** given observed **Actions** and the **Environment**, defined above as probabilistic planning in an MDP. $P(\text{Goal}|\text{Environment})$ is the prior probability of the **Goal** given the **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 **Actions** and the **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 **Actions**, given a **Goal** and the **Environment**, weighted by the posterior over the **Goal** given previously

observed **Actions** and the **Environment** from Equation 2.1. This is just the posterior predictive distribution over future **Actions'** given previous **Actions** and the **Environment**:

$$P(\text{Actions}'|\text{Actions}, \text{Environment}) = \sum_{\text{Goal}} P(\text{Actions}'|\text{Goal}, \text{Environment})P(\text{Goal}|\text{Actions}, \text{Environment}). \quad (2.2)$$

We use Equations 2.1 and 2.2 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 2.1. In Experiment 2, we model people's retrospective goal inferences using a smoothed version of Equation 2.1. In Experiment 3, we model people's predictions of agents' future actions, given observations of their past behavior of varying complexity, using Equation 2.2.

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 2.1 and 2.2 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.

2.2.1 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.

2.3 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.

2.3.1 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. 2-2(a).

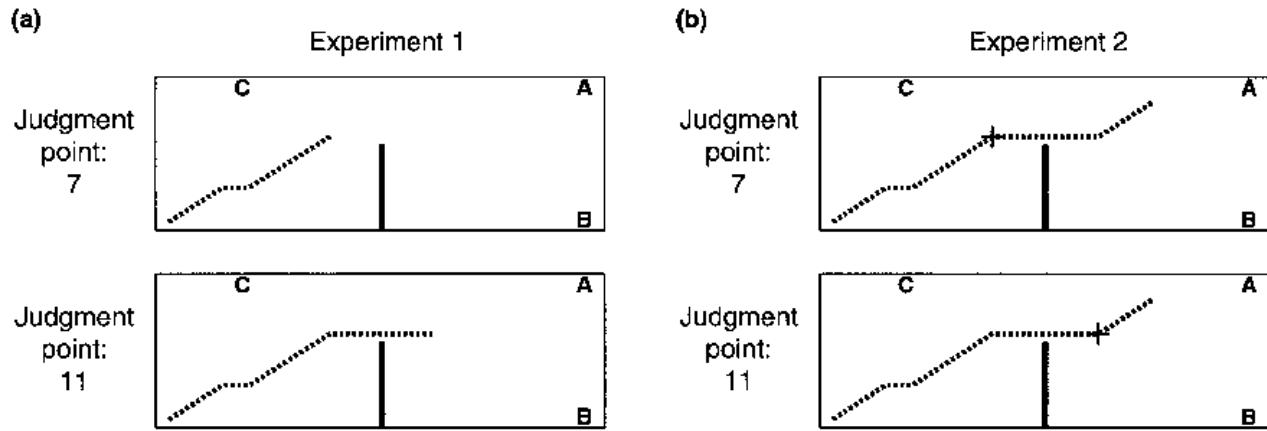


Figure 2-2: Stimulus paradigm for Experiments 1 and 2. Each stimulus presented an animation of an agent's path (marked by a dashed line) ending at a judgment point: a pause in the animation that allowed participants to report their online inferences of the agent's goal at that point. (a) Experiment 1: 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.

Design

All 36 conditions of Experiment 1 are shown in Fig. 2-3. 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. 2-3. 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 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. 2-3 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. 2-3, 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. 2-3.

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. 2-2(a) displays two stimuli with judgment points of 7 and 11, respectively, as they were plotted for our subjects. In Fig. 2-3, 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. 2-3 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. 2-3.

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 instructed that aliens could not pass through walls, but that they could pass through gaps in walls. They were told that after each video,

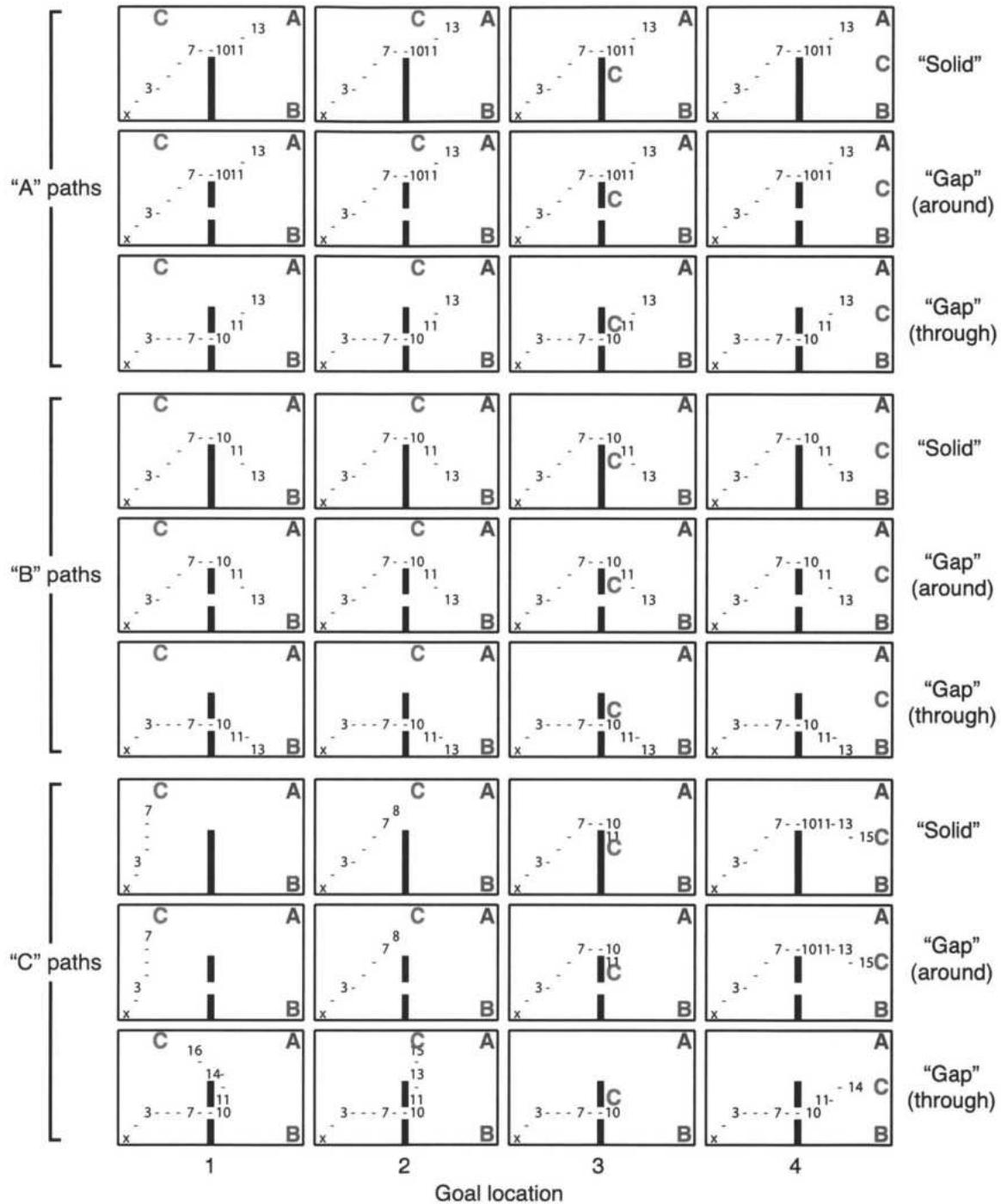


Figure 2-3: All stimuli from Experiment 1. We varied three factors: goal configuration, obstacle shape and agent path. There were four goal configurations, displayed in columns 1-4. Path conditions are grouped as “A” paths, “B” paths and “C” paths. There were two obstacle shape conditions: “Solid” and “Gap”. There were 36 conditions, and 99 unique stimuli in total.

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 2.1 (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.

2.3.2 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. 2-4 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. 2-4(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. 2-4(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.

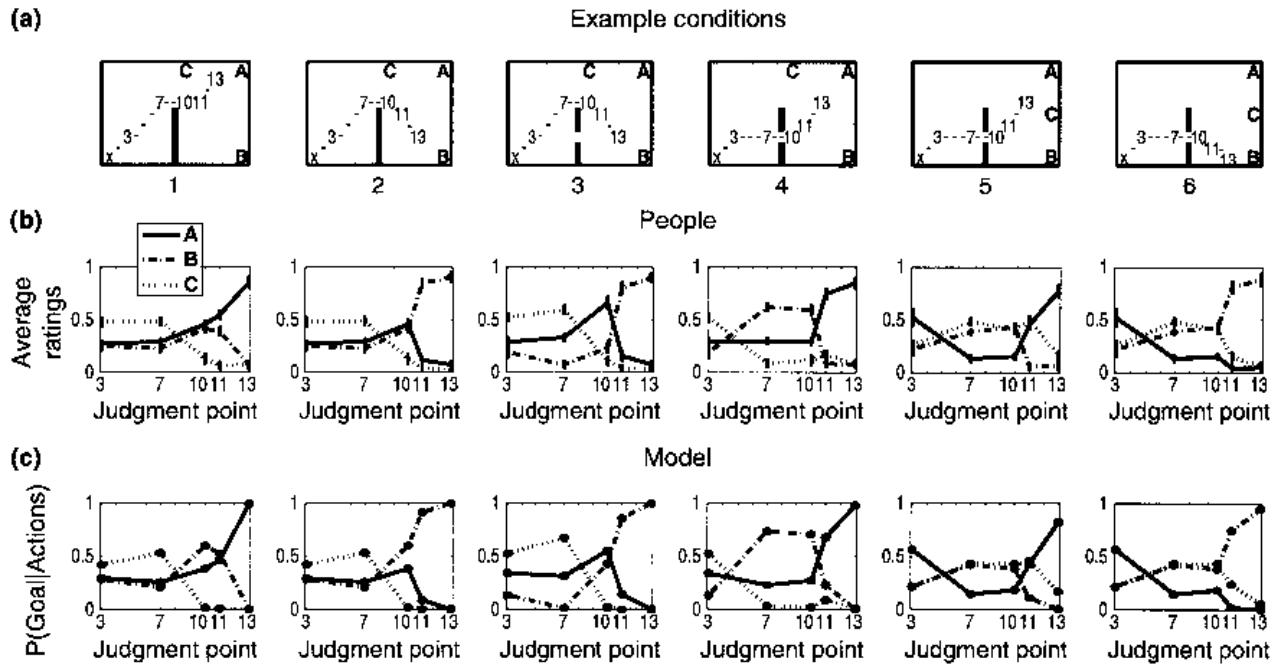


Figure 2-4: Example conditions, data and model predictions from Experiment 1. (a) Stimuli illustrating a range of conditions from the experiment. Each condition differs from adjacent conditions by one feature—either the agent's path, the obstacle, or the goal locations. (b) Average subject ratings with standard error bars for the above stimuli. (c) Predictions of inverse planning model M2 with parameters $\beta = 2.0$, $\gamma = 0.25$.

Participants' average ratings with standard error bars from these conditions are shown in Fig. 2-4(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. 2-4(b)). A comparison of adjacent conditions in Fig. 2-4(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. 2-4(c). The model 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. 2-4 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. 2-4 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. 2-4 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. 2-4 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 2.1. 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 2.1: 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. 2-5 shows scatter plots of the correspondence between participants' judgments and M1, M2, M3 and H using their best-fitting parameter values. In Fig. 2-5, M2 has the fewest outliers. M1 has the most outliers, and M3 and H have a significant number of outliers as well.

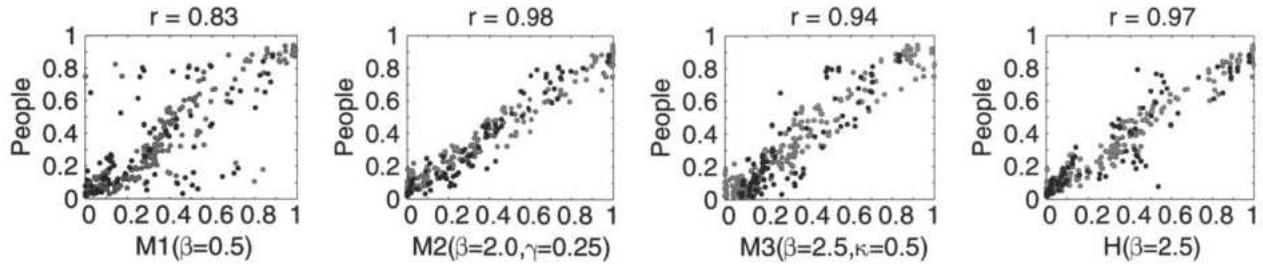


Figure 2-5: Scatter plots of model predictions using best-fitting parameter settings (X-axes) versus people’s online goal inferences (Y-axes) for all Experiment 1 stimuli. Points from the Gap obstacle condition at judgment points 10 and 11 are plotted in black. These trials account for most of the outliers of M1, M3 and H.

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 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 be-

tween 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. 2-6(a), with judgment points 10 and 11 circled. In Fig. 2-5, 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. 2-6, 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. 2-6, 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. 2-6, 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.

2.3.3 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 par-

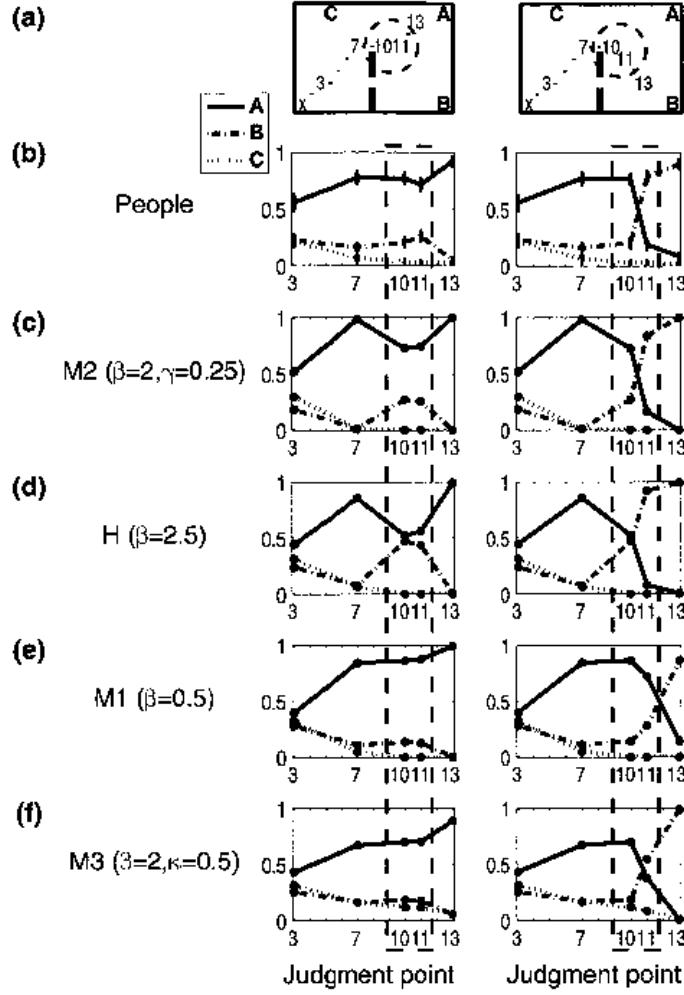


Figure 2-6: Targeted analysis of Experiment 1. (a) Example conditions from our targeted analysis. (b) Participants' ratings from the above conditions. (c) Predictions of M2 ($\beta = 2.0, \gamma = 0.25$). (d) Predictions of H ($\beta = 2.5$). (e) Predictions of M1 ($\beta = 0.5$). (f) Predictions of M3 ($\beta = 2.0, \kappa = 0.5$).

ticipants' 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).

2.4 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.

2.4.1 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. 2-2(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. 2-3. 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 2.1 (specific versions for M1, M2, M3 and H are provided in the appendix).

2.4.2 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. 2-7 shows qualitative comparisons between participants’ goal inferences and the predictions of M2. There is a direct correspondence between the conditions shown in Fig. 2-7 and the conditions in Fig. 2-4 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. 2-7(a) shows retrospective versions of the corresponding conditions in Fig. 2-4(a). Each condition in Fig. 2-7(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.

Fig. 2-7(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 uncer-

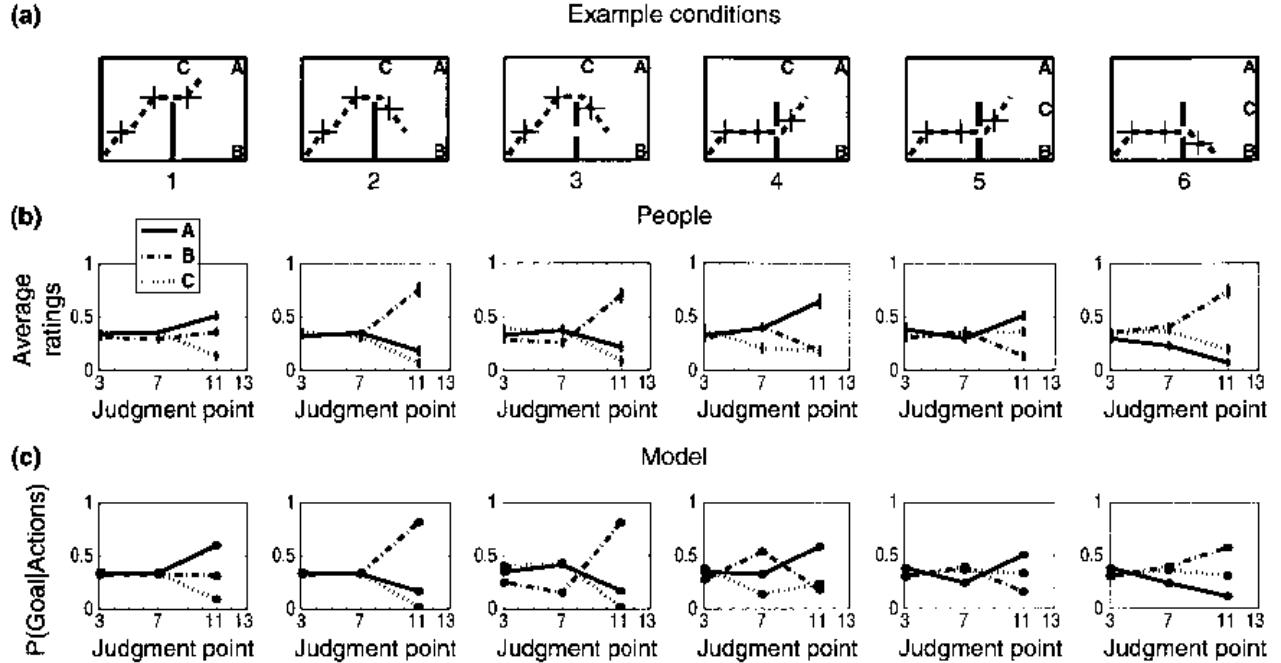


Figure 2-7: Example conditions, data and model predictions from Experiment 2. (a) Stimuli illustrating a range of conditions from the experiment. These stimuli directly correspond to the Experiment 1 stimuli in Fig. 2-4(a). Dashed lines correspond to the movement subjects saw prior to rating the likelihood 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 $\beta = 0.5$, $\gamma = 0.65$.

tainty 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. 2-7(c) shows the predictions of M2 with the best-fitting parameter values. Comparing Fig. 2-7(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 2.2. 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 2.2: 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. 2-8 to illustrate the pattern of errors for each model. M2 predicted people's judgments most accurately, and in Fig. 2-8, 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. 2-8.

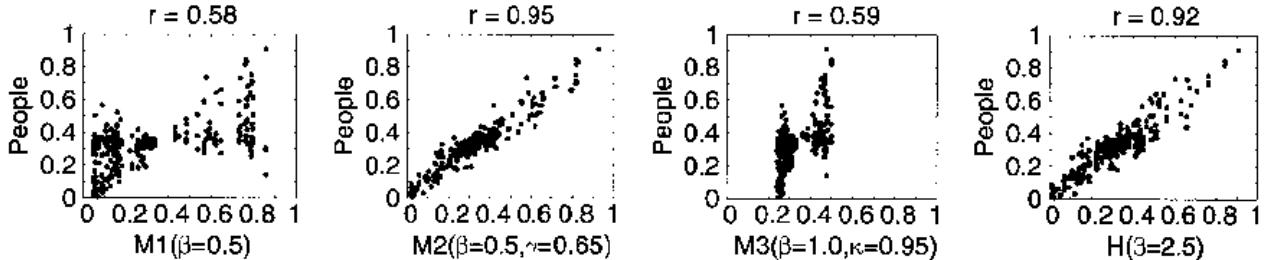


Figure 2-8: Scatter plots of model predictions using best-fitting parameter values (X-axes) versus people's retrospective goal inferences (Y-axes) for all Experiment 2 stimuli.

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.

2.4.3 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.

2.5 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 2.2).

2.5.1 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. 2-9 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. 2-9 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

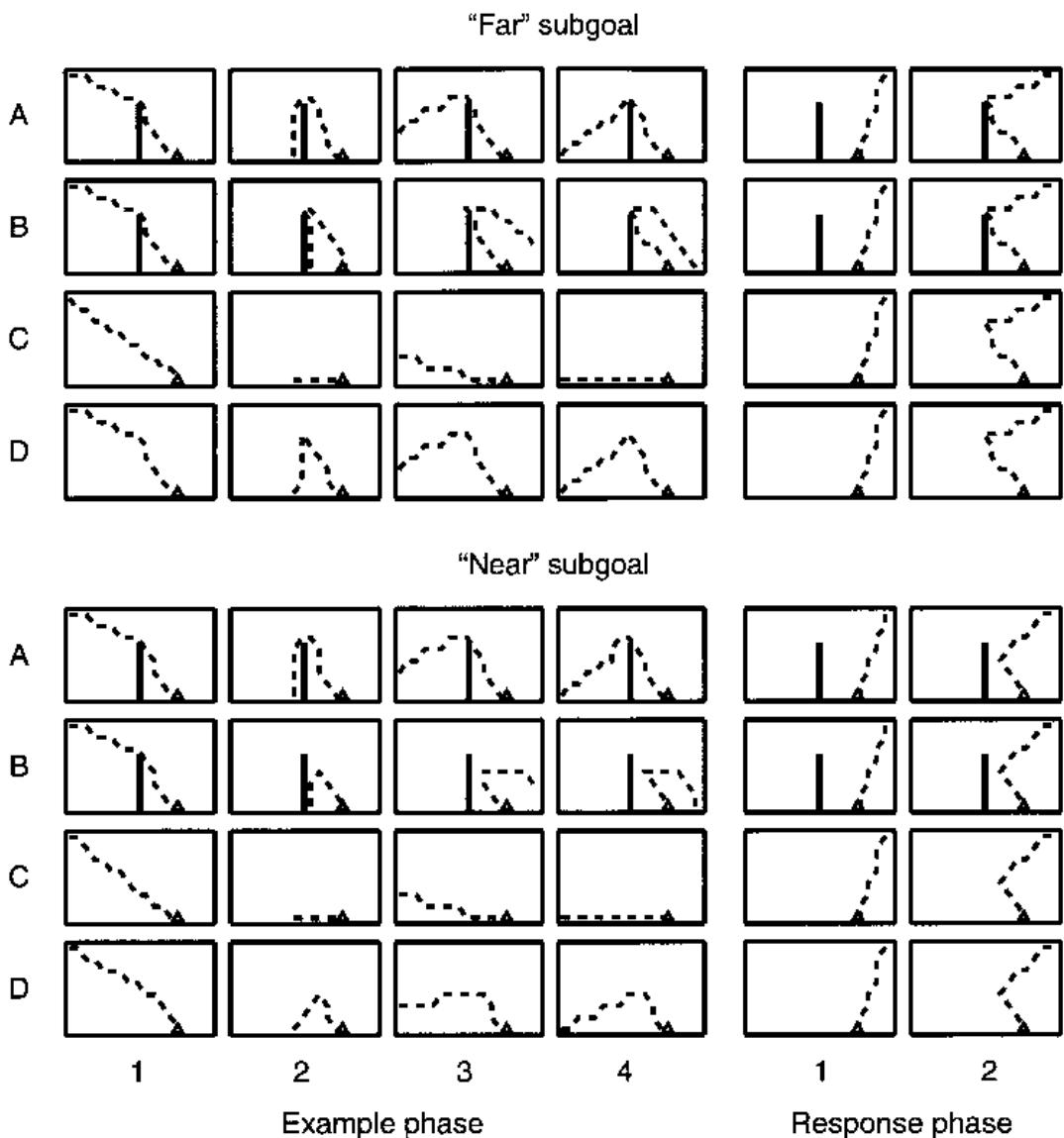


Figure 2-9: All stimuli from Experiment 3. The experiment followed a $2 \times 2 \times 2$ design, which varied the directness of the agent's path, the presence or absence of an obstacle, and the location of potential subgoals. Each condition had 4 trials, which each presented a different example path,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

Fig. 2-9. 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 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 2.2. 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.

2.5.2 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. 2-10 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.

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

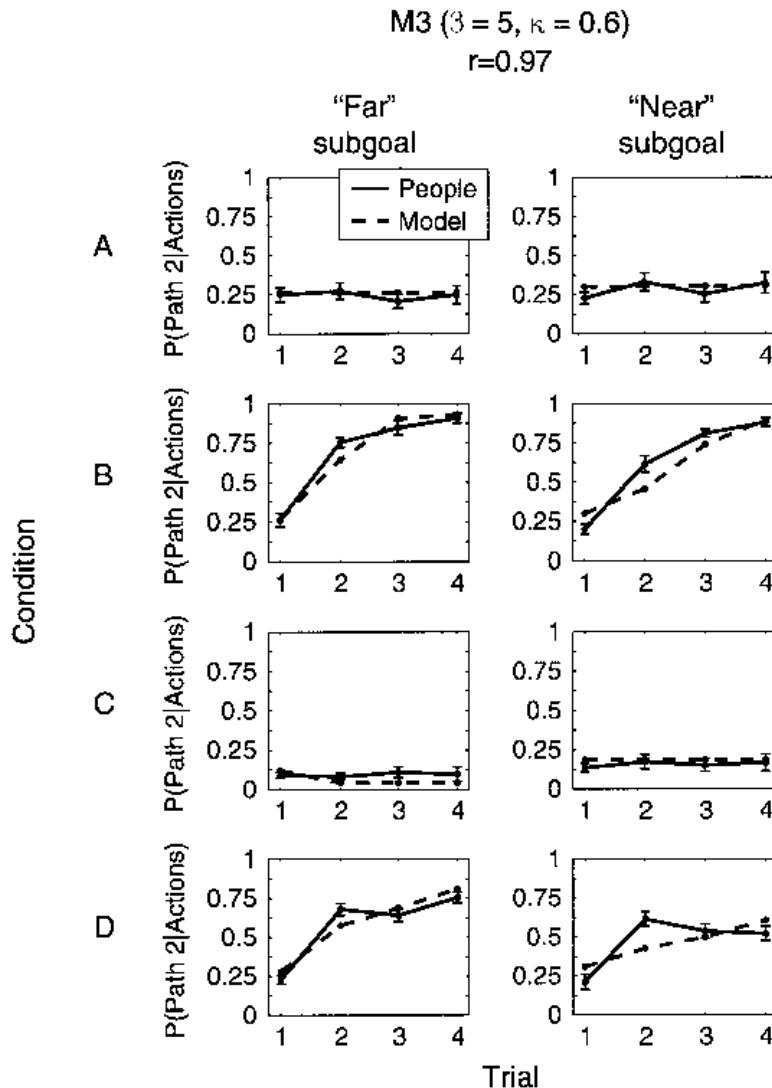


Figure 2-10: Subjects versus M3 with the best-fitting parameters for all stimuli from Experiment 3. Ratings correspond directly to conditions from Fig. 2-9.

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.

2.6 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, Littman, & Cassandra, 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; Kording, 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, Slaughter, & Carey, 1998; Guajardo & Woodward, 2004; Saxe, Tenenbaum, & Carey, 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, Carlin, Stern, & Rubin, 2003; Tenenbaum et al., 2006; Kemp, Perfors, & Tenenbaum, 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 partic-

ular 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 $\log P(\text{Stimuli}|\text{Model})$, 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 2.3. 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.

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

	log $P(\text{Stimuli} \text{Model})$			
	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 2.3: 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.

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, Koller, Parr, & Venkataraman, 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. 2-1(c). Here we focused on the special case of Fig. 2-1(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; A. I. 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.

2.7 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, Knill, & Kersten, 1995; Weiss, Simoncelli, & Adelson, 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, Slovic, & Tversky, 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, Caramazza, & Green, 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.

Chapter 3

Bayesian Theory of Mind

Abstract

We present a computational framework for understanding *Theory of Mind (ToM)*: the human capacity for reasoning about agents' mental states such as beliefs and desires. Our Bayesian model of ToM (or BToM) expresses the predictive model of belief- and desire-dependent action at the heart of ToM as a *partially observable Markov decision process (POMDP)*, and reconstructs an agent's joint belief state and reward function using Bayesian inference, conditioned on observations of the agent's behavior in some environmental context. We test BToM by showing participants sequences of agents moving in simple spatial scenarios and asking for joint inferences about the agents' desires and beliefs about unobserved aspects of the environment. BToM performs substantially better than two simpler variants: one in which desires are inferred without reference to an agent's beliefs, and another in which beliefs are inferred without reference to the agent's dynamic observations in the environment.

3.1 Introduction

Central to human social behavior is a *theory of mind (ToM)*, the capacity to explain and predict people's observable actions in terms of unobservable mental states such as beliefs and desires. Consider the case of Harold, who leaves his dorm room one Sunday morning for the campus library. When he reaches to open the library's front door he will find that it is locked – closed on Sunday. How can we explain his behavior? It seems plausible that he wants to get a book, that he believes the book he wants is at the library, and that he also believes (falsely, it turns out) that the

library is open on Sunday.

Such mental state explanations for behavior go well beyond the observable data, leading to an inference problem that is fundamentally ill-posed. Many different combinations of beliefs and desires could explain the same behavior, with inferences about the strengths of beliefs and desires trading off against each other, and relative probabilities modulated heavily by context. Perhaps Harold is almost positive that the library will be closed, but he needs a certain book so badly that he still is willing to go all the way across campus on the off chance it will be open. This explanation seems more probable if Harold shows up to find the library locked on Saturday at midnight, as opposed to noon on Tuesday. If he arrives after hours already holding a book with a due date of tomorrow, it is plausible that he knows the library is closed and is seeking not to get a new book, but merely to return a book checked out previously to the night drop box.

Several authors have recently proposed models for how people infer others' goals or preferences as a kind of Bayesian inverse planning or inverse decision theory (Baker et al., 2009; Feldman & Tremoulet, 2008; Lucas et al., 2009; Bergen et al., 2010; Yoshida et al., 2008; Ullman et al., 2009). These models adapt tools from control theory, econometrics and game theory to formalize the *principle of rational action* at the heart of children and adults' concept of intentional agency (Gergely et al., 1995; Dennett, 1987): all else being equal, agents are expected to choose actions that achieve their desires as effectively and efficiently as possible, i.e., to maximize their expected utility. Goals or preferences are then inferred based on which objective or utility function the observed actions maximize most directly.

ToM transcends knowledge of intentional agents' goals and preferences by incorporating representational mental states such as subjective beliefs about the world (Perner, 1991). In particular, the ability to reason about *false* beliefs has been used to distinguish ToM from non-representational theories of intentional action (Wimmer & Perner, 1983; Onishi & Baillargeon, 2005). Our goal in this paper is to model human ToM within a Bayesian framework. Inspired by models of inverse planning, we cast Bayesian ToM (BToM) as a problem of inverse planning and inference, representing an agent's planning and inference about the world as a partially observable Markov decision process (POMDP), and inverting this forward model using Bayesian inference. Criti-

cally, this model includes representations of both the agent’s desires (as a utility function), and the agent’s own subjective beliefs about the environment (as a probability distribution), which may be uncertain and may differ from reality. We test the predictions of this model quantitatively in an experiment where people must simultaneously judge beliefs and desires for agents moving in simple spatial environments under incomplete or imperfect knowledge.

Important precursors to our work are several computational models (Goodman et al., 2006; Bello & Cassimatis, 2006a; Goodman et al., 2009) and informal theoretical proposals by developmental psychologists (Wellman, 1990; Gopnik & Meltzoff, 1997; Gergely & Csibra, 2003). Goodman et al. (2006) model how belief and desire inferences interact in the classic “false belief” task used to assess ToM reasoning in children (Wimmer & Perner, 1983). This model instantiates the schema shown in Fig. 3-1(a) as a causal Bayesian network with several psychologically interpretable, but task-dependent parameters. Goodman et al. (2009) model adult inferences of an agent’s knowledge of the causal structure of a simple device (“Bob’s box”) based on observing the agent interacting with the device. To our knowledge, our work here is the first attempt to explain people’s joint inferences about agents’ beliefs and desires by explicitly inverting POMDPs – and the first model capable of reasoning about the graded strengths and interactions between agents’ beliefs and desires, along with the origins of agents’ beliefs via environmentally constrained perceptual observations.

3.2 Computational Framework

This section describes Bayesian Theory of Mind (BToM): a theory-based Bayesian framework (Tenenbaum et al., 2006) that characterizes ToM in terms of Bayesian inference over a formal, probabilistic version of the schema in Fig. 3-1(a). BToM represents an ideal *observer* using a theory of mind to understand the actions of an individual *agent* within some environmental context. This ideal-observer analysis of ToM asks how closely human judgments approach the ideal limit, but also what mental representations are necessary to explain human judgments under hypothetically unbounded computational resources. We will first describe BToM in general, but informal terms before progressing to the mathematical details involved in modeling our experimental domain.

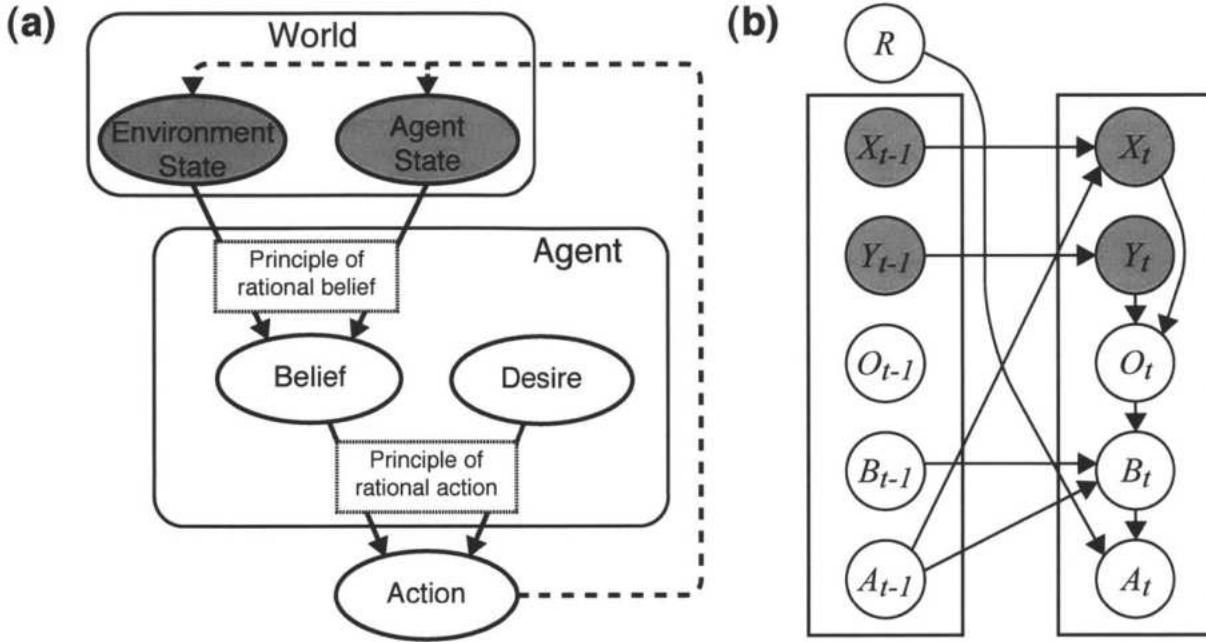


Figure 3-1: Causal structure of theory of mind. Grey shaded nodes are assumed to be observed (for the observer; not necessarily for the agent, as described in the main text). **(a)** Schematic model of theory of mind. Traditional accounts of ToM (e.g. Dennett, 1987; Wellman, 1990; Gopnik & Meltzoff, 1997) have proposed informal versions of this schema, characterizing the content and causal relations of ToM in commonsense terms, e.g., “seeing is believing” for the principle of rational belief. **(b)** Observer’s grounding of the theory as a dynamic Bayes net (DBN). The DBN encodes the observer’s joint distribution over an agent’s beliefs $B_{1:T}$ and desires R over time, given the agent’s physical state sequence $x_{1:T}$ in environment y .

3.2.1 Informal sketch

For concreteness, we use as a running example a simple spatial context (such as a college campus or urban landscape) defined by buildings and perceptually distinct objects, with agents’ actions corresponding to movement, although in general BTOM can be defined over arbitrary state and action spaces (for example, a card game where the state describes players’ hands and actions include draw or fold). The observer’s representation of the world is composed of the *environment state* and the *agent state* (Fig. 3-1(a)). In a spatial context, the state of the environment represents its physical configuration, e.g., the location of buildings and objects, and the state of the agent specifies its objective, external properties, such as its physical location in space.

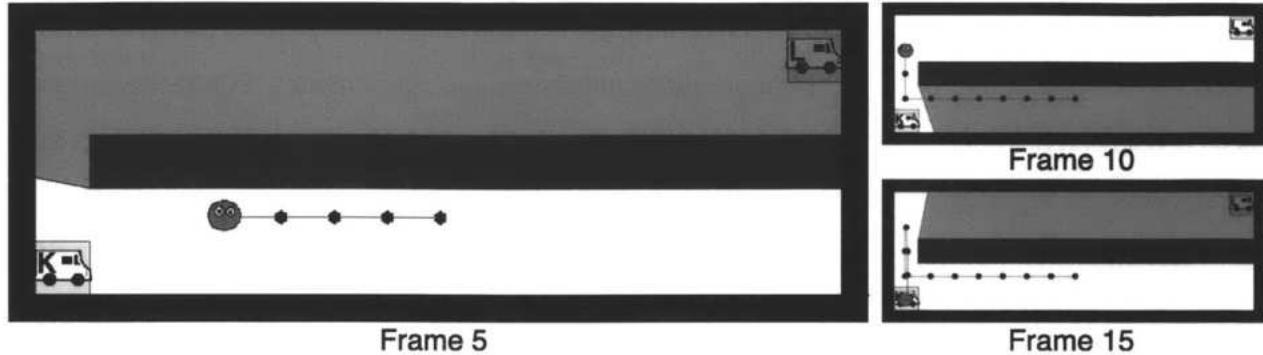


Figure 3-2: Example experimental stimulus. The small blue sprite represents the location of the agent, and the black trail with arrows superimposed records the agent’s movement history. The two yellow cells in opposite corners of the environment represent spots where trucks can park, and each contains a different truck. The shaded grey area of each frame represents the area that is outside of the agent’s current view.

The observer’s theory of the agent’s mind includes representations of the agent’s subjective desires and beliefs, and the principles by which desires and beliefs are related to actions and the environment. Similar to previous models, the content of the agent’s desire consists of objects or events in the world. The agent’s degree of desire is represented in terms of the subjective reward received for taking actions in certain states, e.g., acting to attain a goal while in close proximity to the goal object. The agent can also act to change its own state or the environment state at a certain cost, e.g., navigating to reach a goal may incur a small cost at each step.

The main novel component of the current model is the inclusion of a representation of beliefs. Like desires, beliefs are defined by both their content and the strength or degree with which they are held. The content of a belief is a representation corresponding to a possible world. For instance, if the agent is unsure about the location of a particular object, its belief contents are worlds in which the object is in different locations. The agent’s degree of belief reflects the subjective probability it assigns to each possible world.

The principles governing the relation between the world and the agent’s beliefs, desires and actions can be naturally expressed within partially observable Markov decision processes (POMDPs). POMDPs capture the causal relation between beliefs and the world via the principle of rational belief, which formalizes how the agent’s belief is affected by observations in terms of Bayesian belief updating. Given an observation, the agent updates its degree of belief in a particular world based

on the likelihood of receiving that observation in that world. In a spatial setting, observations depend on the agent’s line-of-sight visual access to features of the environment. POMDPs represent how beliefs and desires cause actions via the principle of rational action, or rational planning. Intuitively, rational POMDP planning provides a predictive model of an agent optimizing the tradeoff between exploring the environment to discover the greatest rewards, and exploiting known rewards to minimize costs incurred.

On observing an agent’s behavior within an environment, the beliefs and desires that caused the agent to generate this behavior are inferred using Bayesian inference. The observer maintains a hypothesis space of joint beliefs and desires, which represent the agent’s initial beliefs about the environment state and the agent’s static desires for different goals. For each hypothesis, the observer evaluates the likelihood of generating the observed behavior given the hypothesized belief and desire. The observer integrates this likelihood with the prior over mental states to infer the agent’s joint belief and desire.

As an example of how this works, consider Fig. 3-2. The “college campus” environment is characterized by the campus size, the location and size of buildings, and the location of several different goal objects, here “food trucks”. The agent is a hungry graduate student, leaving his office and walking around campus in search of satisfying lunch food. There are three trucks that visit campus: Korean (K), Lebanese (L) and Mexican (M), but only two parking spots where trucks are allowed to park, highlighted with a yellow background in Fig. 3-2. The student’s field of view is represented by the unshaded region of the environment.

In Fig. 3-2, the student can initially only see where K (but not L) is parked. Because the student can see K, they know that the spot behind the building either holds L, M, or is empty. By frame 10, the student has passed K, indicating that they either want L or M (or both), and believe that their desired truck is likely to be behind the building (or else they would have gone straight to K under the principle of rational action). After frame 10, the agent discovers that L is behind the building and turns back to K. Obviously, the agent prefers K to L, but more subtly, it also seems likely that the agent wants M more than either K or L, despite M being absent from the scene! BToM captures this inference by resolving the desire for L or M over K in favor of M after the agent rejects L.

In other words, BToM infers the best explanation for the observed behavior – the only consistent desire that could lead the agent to act the way it did.

3.2.2 Formal modeling

In the food-truck domain, the agent occupies a discrete state space \mathcal{X} of points in a 2D grid. The environment state \mathcal{Y} is the set of possible assignments of the K, L and M trucks to parking spots. Possible actions include North, South, East, West, Stay, and Eat. Valid actions yield the intended transition with probability $1-\epsilon$ and do nothing otherwise; invalid actions (e.g., moving into walls) have no effect on the state.

The agent’s visual observations are represented by the *isovist* from the agent’s location: a polygonal region containing all points of the environment within a 360-degree field of view (Davis & Benedikt, 1979; Morariu, Prasad, & Davis, 2007). Example isovists from different locations in one environment are shown in Fig. 3-2. The observation distribution $P(o|x, y)$ encodes which environments in \mathcal{Y} are consistent with the contents of the isovist from location x . We model observation noise with the simple assumption that ambiguous observations can occur with probability ν , as if the agent failed to notice something that should otherwise be visible.

The observer represents the agent’s belief as a probability distribution over \mathcal{Y} ; for $y \in \mathcal{Y}$, $b(y)$ denotes the agent’s degree of belief that y is the true state of the environment. Bayesian belief updating at time t is a deterministic function of the prior belief b_{t-1} , the observation o_t , and the world state (x_t, y) . The agent’s updated degree of belief in environment y satisfies $b_t(y) \propto P(o_t|x_t, y)b_{t-1}(y)$.

The agent’s reward function $R(x, y, a)$ encodes the subjective utility the agent derives from taking action a from the state (x_t, y) . Each action is assumed to incur a cost of 1. Rewards result from taking the “Eat” action while at a food truck; the magnitude of the reward depends on the strength of the agent’s desire to eat at that particular truck. Once the student has eaten, all rewards and costs cease, implying that rational agents should optimize the tradeoff between the number of actions taken and the reward obtained.

The agent’s POMDP is defined by the state space, the action space, the world dynamics, the

observation model, and the reward function. We approximate the optimal value function of the POMDP for each hypothesized reward function using a point-based value iteration algorithm over a uniform discretization of the belief space. The agent’s policy is stochastic, given by the softmax of the lookahead state-action value function Q^{LA} (Hauskrecht, 2000): $P(a|b, x, y) \propto \exp(\beta Q^{LA}(b, x, y, a))$. The β parameter establishes the degree of determinism with which the agent executes its policy, capturing the intuition that agents tend to, but do not always follow the optimal policy.

Our approach to joint belief and desire inference is closely related the model of belief filtering in Zettlemoyer, Milch, and Kaelbling (2009), restricted to the case of one agent reasoning about the beliefs of another. Fig. 3-1(b) shows the observer’s dynamic Bayes net (DBN) model of an agent’s desires, states, observations, beliefs and actions over time. The observer’s belief and reward inferences are given by the joint posterior marginal over the agent’s beliefs and rewards at time t , given the state sequence up until $T \geq t$: $P(b_t, r|x_{1:T}, y)$. This computation is analogous to the forward-backward algorithm in hidden Markov models, and provides the basis for model predictions of people’s joint belief and desire inferences in our experiment.

To perform inference over the multidimensional, continuous space of beliefs and rewards, we uniformly discretize the hypothesis spaces of beliefs and reward functions with grid resolutions of 7. The range of reward values was calibrated to the spatial scale of our environments, taking values $-20, 0, \dots, 100$ for each truck. Model predictions were based on the student’s expected reward value for each truck (K, L, M) and the expected degree-of-belief in each possible world for each trial.

3.2.3 Alternative models

To test whether the full representational capacity of our model is necessary to understand people’s mental state attributions, we formulate two alternative models as special cases of our joint inference model. Each alternative model “lesions” a central component of the full model’s representation of beliefs, and tests whether it is possible to explain people’s inferences about agents’ desires in our experiment without appeal to a full-fledged theory of mind.

Our first alternative model is called TrueBel. This model assumes that the state is fully observable to the agent, i.e., that the agent knows the location of every truck, and plans to go directly to the truck that will provide the maximal reward while incurring the least cost. We hypothesized that this model would correlate moderately well with people's desire judgments, because of the statistical association between desired objects and actions.

Our second alternative model is called NoObs. In this model, the agent has an initial belief about the state of the environment, but there is no belief updating – the initially sampled belief remains fixed throughout the trial. We hypothesized that this model might fit people's belief and desire inferences in situations where the agent appeared to move toward the same truck throughout the entire trial, but that for actions that required belief updating or exploration to explain, for instance, when the agent began by exploring the world, then changed direction based on its observation of the world state, NoObs would fit poorly.

3.3 Experiment

Fig. 3-4 illustrates our experimental design. Truck labels were randomized in each trial of the experiment, but we will describe the experiment and results using the canonical, unscrambled ordering Korean (K), Lebanese (L), Mexican (M).

The experiment followed a $3 \times 5 \times 2 \times 3 \times 2$ design. These factors can be divided into 30 ($3 \times 5 \times 2$) unique paths and 6 (3×2) unique environmental contexts. There were 3 different starting points in the environment: "Left", "Middle", or "Right"; all shown in Fig. 3-4. These starting points were crossed with 5 different trajectories: "Check-Left, go to K"; "Check-Left, go to L/M"; "Check-Right, go to K"; "Check-Right, go to L/M"; and "No-check, go straight to K". Four of these trajectories are shown in Fig. 3-4. Each path was shown with 2 different judgment points, or frames at which the animation paused and subjects gave ratings based on the information shown so far. Judgment points were either at the moment the student became able to see the parking spot that was initially occluded ("Middle"; e.g., frame 10 in Fig. 3-2), or at the end of the path once the student had eaten ("Ending"; e.g., frame 15 in Fig. 3-2). All potential paths were crossed with 6 environmental contexts, generated by combining 3 different building configurations: "O", "C" and

“backwards C”, (all shown in Fig. 3-4) with 2 different goal configurations: “One truck” or “Two trucks” present; both shown in Fig. 3-4.

After all possible trials from this design were generated, all invalid trials (in which the student’s path intersected with a building), and all “Ending” trials in which the path did not finish at a truck were removed. This left 78 total trials. Of these, 5 trials had a special status. These were trials in the “O” environment with paths in which the student began at the Right starting point, and then followed a Check-Left trajectory. These paths had no rational interpretation under the BToM model, because the Check-Right trajectory was always a more efficient choice, no matter what the student’s initial belief or desire. These “irrational” trials are analyzed separately in the Results section.

Several factors were counterbalanced or randomized. Stimulus trials were presented in pseudo-random order. Each trial randomly scrambled the truck labels, and randomly reflected the display vertically and horizontally so that subjects would remain engaged with the task and not lapse into a repetitive strategy. Each trial randomly displayed the agent in 1 of 10 colors, and sampled a random male or female name without replacement. This ensured that subjects did not generalize information about one student’s beliefs or desires to students in subsequent trials.

The experimental task involved rating the student’s degree of belief in each possible world (Lebanese truck behind the building (L); Mexican truck behind the building (M); or nothing behind the building (N)), and rating how much the student liked each truck. All ratings were on a 7-point scale. Belief ratings were made retrospectively, meaning that subjects were asked to rate what the student thought was in the occluded parking spot before they set off along their path, basing their inference on the information from the rest of the student’s path. The rating task counterbalanced the side of the monitor on which the “likes” and “believes” questions were displayed.

Subjects first completed a familiarization stage that explained all details of our displays and the scenarios they depicted. To ensure that subjects understood what the students could and couldn’t see, the familiarization explained the visualization of the student’s isovist, which was updated along each step of the student’s path. The isovist was displayed during the testing stage of the experiment as well (Fig. 3-2).

Participants were 17 members of the MIT subject pool, 6 female, and 11 male. One male subject did not understand the instructions and was excluded from the analysis.

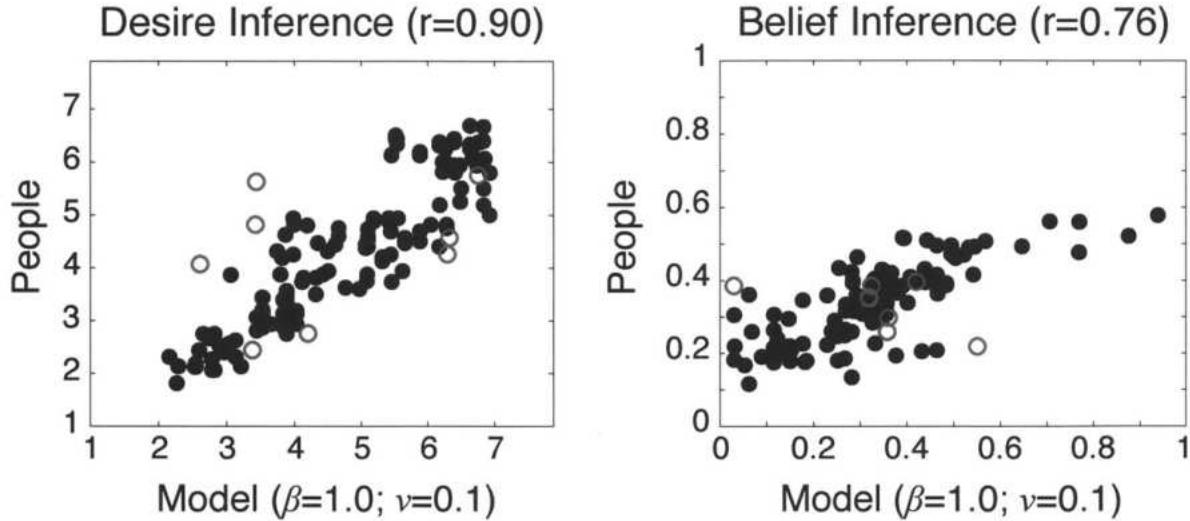
3.3.1 Results & Discussion

Debriefing of subjects suggested that many were confused by the “Middle” judgment point trials; this was also reflected by greater variability in people’s judgments within these trials. Because of this, our analyses only include trials from the “Ending” judgment point condition, which accounted for 54 out of the 78 total trials.

We begin by analyzing the overall fit between people’s judgments and our three models, and then turn to a more detailed look at several representative scenarios. Two parameters β and ν were fit for the BToM model; only the determinism parameter β is relevant for the TrueBel and NoObs models. Parameter fits are not meant to be precise; we report the best values found among several drawn from a coarse grid.

BToM predicts people’s judgments about agents’ desires relatively well, and less well but still reasonably for judgments about agents’ initial beliefs (Fig. 3-3). In Fig. 3-3, data from the “irrational” trials are plotted with magenta circles, and account for most of the largest outliers. TrueBel and NoObs fit significantly worse for desire judgments and provide no reasonable account of belief judgments. TrueBel’s belief predictions are based on the actual state of the world in each trial; the poor correlation with people’s judgments demonstrates that people did not simply refer to the true world state in their belief attributions. The NoObs model in principle can infer agents’ beliefs, but without a theory of how beliefs are updated from observations it must posit highly implausible initial beliefs that correlate poorly with subjects’ judgments over the whole set of experimental conditions.

Fig. 3-4 shows several revealing comparisons of human judgments and model predictions in specific cases. When the agent follows a long path to an unseen goal (A1) it is suggestive of a strong initial belief that a more desirable truck is present behind the wall. In contrast, going straight to a nearby observed truck says only that this truck is likely to be desired more than the others (A2). When the agent goes out of its way to check an unseen parking spot, sees the second



	BToM	TrueBel	NoObs
Desire	0.90	0.67	0.61
Belief	0.76	0.11	0.39

Figure 3-3: Scatter plots show overall correlations between BToM model predictions and human judgments about desires and beliefs in our experiment. Each dot corresponds to the mean judgment of subjects in one experimental condition. Magenta circles correspond to trials which had no rational interpretation in terms of POMDP planning. The table shows correlations with human judgments for BToM and two simpler variants, which do not represent beliefs (TrueBel) or do not update beliefs based on observations (NoObs).

truck there, and returns to the previously seen truck, it suggests a strong desire for the one truck not present (compare B1 to B2). Finally, the relative strengths of inferences about desires and initial beliefs are modulated by how far the agent must travel to observe the unseen parking spot (compare C1 to C2, and C3 to C4). In each of these cases people reflect the same qualitative trends predicted by the model.

The finding that people's inferences about agents' desires are more robust than inferences about beliefs, and more consistent with the model's predictions, is intriguingly consistent with classic asymmetries between these two kinds of mental state attributions in the ToM literature. Intentional actions are the joint consequence of an agent's beliefs and desires, but inferences from actions back to beliefs will frequently be more difficult and indirect than inferences about desires. Actions often point with salient perceptual cues directly toward an agent's goal or desired state. When a

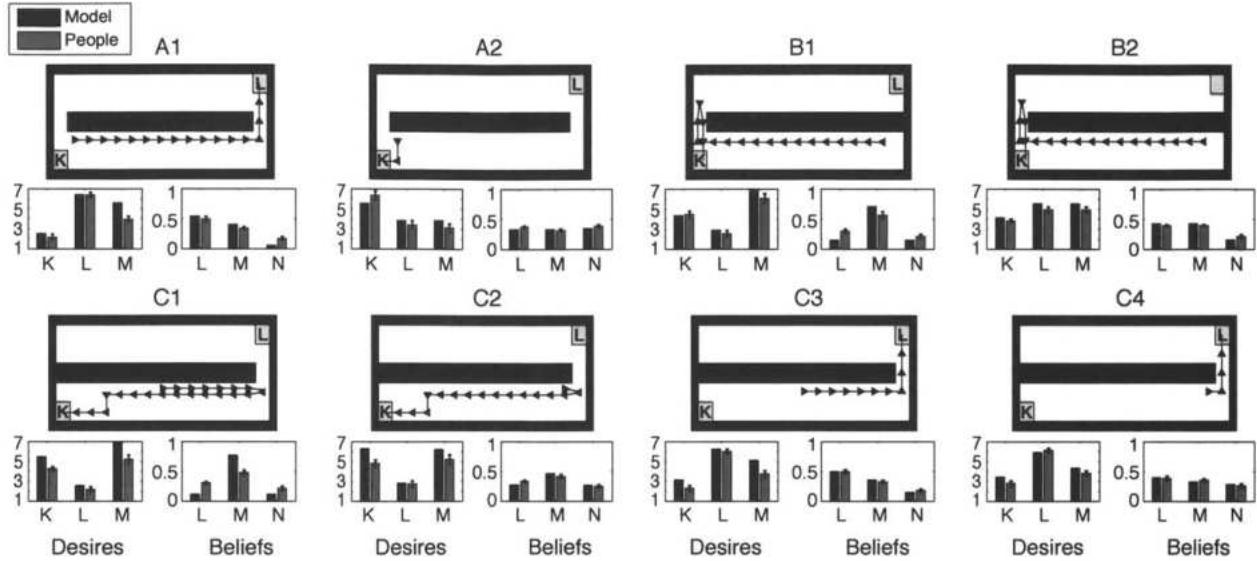


Figure 3-4: Eight representative scenarios from the experiment, showing the agent’s path, BTOM model predictions for the agent’s desires (for trucks K, L or M, on a scale of 1 to 7) and beliefs about the unseen parking spot (for trucks L, M or no truck (N), normalized to a probability scale from 0 to 1), and mean human judgments for these same mental states. Error bars show standard error ($n=16$).

person wants to take a drink, her hand moves clearly *toward* the glass on the table. In contrast, no motion so directly indicates what she believes to be inside the glass. Infants as young as five months can infer agents’ goals from their actions (Gergely & Csibra, 2003), while inferences about representational beliefs seem to be present only in rudimentary forms by age one and a half, and in more robust forms only by age 4 (Onishi & Baillargeon, 2005).

3.4 Conclusion & Future Work

Our experiment showed that human ToM inferences come surprisingly close to those of an ideal rational model, performing Bayesian inference over beliefs and desires simultaneously. By comparing with two alternative models we showed that it was necessary to perform joint inference about agents’ beliefs and desires, and to explicitly model the agent’s observational process, as part of modeling people’s theory of mind judgments. Crucially, it was also necessary to represent initial uncertainty over both the agent’s beliefs and desires.

We have not attempted to distinguish here between agents' general desires and their specific goals or intentions at particular moments of action. In previous work we showed that inferences about which object is most likely to be an agent's instantaneous goal were well explained using a similar Bayesian inverse planning framework (Baker et al., 2009). However, goals are not always about objects. In the present experiments, it feels intuitive to describe agents as attempting to maximize their overall expected utility by adopting a combination of object- and information-seeking goals (or goals intended to update the agent's beliefs). For instance, in Fig. 3-4, B1 it looks as if the agent initially had a goal of finding out which truck was parked on the other side of the wall, and then after failing to find their preferred truck (M) there, set a goal of returning to the previously observed second-favorite truck (K). Our model can produce and interpret such behavior, but it does so without positing these explicit subgoals or the corresponding parse of the agent's motion into subsequences, each aimed to achieve a specific goal. Extending our model to incorporate a useful intermediate representation of goal sequences is an important direction for future work. Even without these complexities, however, we find it encouraging to see how well we can capture people's joint attributions of beliefs and desires as Bayesian inferences over a simple model of rational agents' planning and belief updating processes.

Chapter 4

Theory-based Social Goal Inference

Abstract

Everyday human interaction relies on making inferences about social goals: goals that an intentional agent adopts in relation to another agent, such as “approaching” or “avoiding”. We present a computational model of social goal inference that takes as input observations of multiple agents moving in some environmental context. The model infers a social goal for each agent that is most likely to have given rise to that agent’s observed actions, under an intuitive theory that expects agents to act approximately rationally. We compare this theory-based approach to a simpler bottom-up cue-based approach in a behavioral experiment designed to distinguish these accounts. Finally, we consider the need to incorporate second-order reasoning in our approach, to model people’s reasoning about agents reasoning about other agents’ goals in formulating their action plans.

4.1 Introduction

Even the simplest everyday social interactions are supported by complex mental representations and processes. Contingent on the environment, our prior knowledge and our desires, we form goals, beliefs, intentions and other mental states with rich representational content. When we observe or participate in social interactions, we make joint inferences about the mental states of multiple interacting agents. For instance, watching a short film of two people moving inside a house, we might interpret it as follows: “X is trying to catch Y, and Y is trying to escape. X is sneaking up on Y in the bedroom but doesn’t realize that Y has already seen him and is preparing

to escape out the back.” Inferring such intentional descriptions of social interactions is computationally challenging. Other agents’ mental states are not directly observable, and the space of possible beliefs and goals that one agent could have with respect to another is effectively infinite. Yet these inferences are remarkably quick and robust. Adults (Heider & Simmel, 1944) and even infants (Kuhlmeier et al., 2003; Gergely et al., 1995) can infer relational goals such as “chasing”, “helping”, and “harming” from brief animations of simple shapes moving in a two-dimensional environment.

To account for inferences like these, it has been proposed that people draw on an intuitive theory of psychology, which may already be present in infancy in some simpler form as a rational agent schema (Gopnik & Wellman, 1992; Gopnik & Meltzoff, 1997; Gergely et al., 1995; R. Gelman et al., 1995). This approach casts the interpretation of intentional action as a top-down process, drawing on a causal model of how agents’ beliefs, intentions and environmental and social context jointly influence their behavior. Agents’ mental states are inferred from observed social interactions through a process of “inference to the best explanation”, positing the goals and beliefs under which their observed actions are most plausible in the observed context.

A key challenge for theory-based accounts of action understanding is precisely specifying the nature of the causal relation between intentions, context and behavior. Several authors (Dennett, 1987; Gergely et al., 1995) have argued that this relation embodies the *principle of rationality*: the assumption that intentional agents should act to achieve their goals as efficiently as possible, given their beliefs about the world. The principle of rationality is appealing because it applies very generally across agents, intentions, and contexts, providing a way to dynamically build causal models of agents’ behavior in novel situations. However, the sense of rationality that people apply can be nuanced, particularly in the case of social interaction, where the rational strategy of an agent might depend on its expectations of others’ behavior, whose rational strategies depend on expectations of its behavior, and so on *ad infinitum*.

An alternative approach (Blythe et al., 1999; Barrett et al., 2005; Zacks, 2004) to specifying the relation between mental states, context and behavior emphasizes simple visual cues that can be computed through bottom-up perceptual processes. Movements with a particular social intention,

such as “chasing” or “fleeing”, are perceived as instances of categories defined by prototypical motion patterns. For example, by measuring the relative directions and velocities of two agents’ motions, we might be able to identify if one is trying to catch the other. A simple cue-based approach appeals in part because it seems to be the kind of computation that can be done rapidly, robustly and in young infants without sophisticated cognitive capacities. It can also be formalized precisely in computational models (Blythe et al., 1999). However, it is unlikely to be sufficient to explain human social inference. The inferences of both adults and infants are highly sensitive to the environmental and task contexts in which actions take place, in ways that are hard to explain by a bottom-up motion analysis but suggest a deeper causal basis. For instance, there are situations in which Y may be trying to avoid X, but Y’s most likely path of motion is to head *toward* X rather than away from X; Y may be trapped in a corner with only route of escape that runs directly past X.

Theory-based approaches seem to offer a deeper account of social goal inference, but unlike bottom-up cue-based models, they have not been worked out computationally or tested rigorously. These are the goals our present paper. We give a Bayesian formulation of the theory-based approach to inferring social goals from observed action sequences. Our framework can capture a range of theory-based models, differing in how richly they model agents’ mental states. We show how different theory-based approaches can be tested against each other and against simpler motion cue-based approaches.

Previously (Baker et al., 2006, 2007) we showed how to model action understanding based on the rationality principle as a Bayesian inference about the structure of a goal-based Markov Decision Problem (MDP), a rational model for probabilistic planning. Here, we extend this work to model people’s inferences about social goals, such as approaching and avoiding, from observations of multiple agents interacting in simple environments. People’s representations of other minds might include not just first-order content, such as agents’ intentions or beliefs about the state of the world, but also second-order content: representing agent’s intentions, contingent on *their* representation of other agents. Second-order mentalizing is particularly important in social interaction, where understanding social relations between agents such as helping, loving, loathing,

and empathizing require the ability to represent a mind representing a mind.

The plan of the paper is as follows. We first introduce the computational framework, and then describe a quantitative behavioral experiment and several qualitative phenomena of social goal inference that can be used to distinguish alternative models. We compare the accuracy of both cue-based models and theory-based models in predicting how people categorize and extrapolate the interactions of pairs of agents moving in simple environments. We close by discussing the insights gained into the representations and principles underlying people's mental state inferences.

4.2 Computational framework

Most abstractly, our framework represents the structure of the causal relation between agents' environmental and social context, their mental states, and their behavior. Fig. 4-1 illustrates how models of rational planning allow observers to dynamically generate a conditional probability distribution over agents' actions (A_i), given their goals (G_i), their beliefs about other agents' goals ($B_i(G_j)$; "social context") and the state of the world (W). For simplicity, we consider situations with two agents, who both have complete knowledge of the world state $W = \langle S_1, S_2, E \rangle$, which includes both agent states S_1, S_2 and the environment E . The environment is assumed to be static, while agent states are dynamic. In Fig. 4-1, agents choose their actions independently of each other, so the conditional joint probability over both actions factors into the product of the conditional probability of each agent's action:

$$P(A_1, A_2 | G_1, G_2, B_1(G_2), B_2(G_1), W) = P(A_1 | G_1, B_1(G_2), W)P(A_2 | G_2, B_2(G_1), W). \quad (4.1)$$

Inferring mental states given observed actions and the world state is done by inverting this model using Bayes' rule:

$$\begin{aligned} P(G_1, G_2, B_1(G_2), B_2(G_1) | A_1, A_2, W) &\propto \\ P(A_1 | G_1, B_1(G_2), W)P(A_2 | G_2, B_2(G_1), W)P(G_1, B_1(G_2))P(G_2, B_2(G_1)). \end{aligned} \quad (4.2)$$

Equation 4.2 describes a kind of “action parse”, assigning social goals and beliefs about others’ social goals to all agents by integrating bottom-up information from the observed actions with top-down information from the prior over agents’ mental states, $P(G_i, B_i(G_j))$.

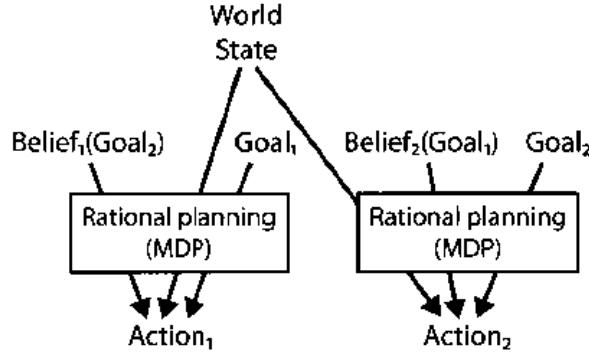


Figure 4-1: Causal schema for an intuitive theory of social interaction (for two agents).

We consider two types of agents: Level 1 (L1) and Level 2 (L2). Agents of both types have social goals, but only L2 agents represent other agents’ goals ($B_i(G_j)$). Without representing others’ goals, L1 agents cannot anticipate the actions of other agents, and we model their predictions as being maximally uncertain, predicting a random walk for other agents. By representing other agents’ social goals, L2 agents can predict agents’ future behavior to pursue their own goals more efficiently.

For concreteness, we consider two social goals for L1 and L2 agents: Approach and Avoid. The meaning of these goals roughly corresponds to the meaning of their words in English; however, their precise semantics is defined by their role in our theory-based models of social interaction. We assume a uniform prior over goals, so that cooperation and competition are equally likely. Next, we describe how Equation 4.1 is computed for these social goals.

4.2.1 (Inverse) planning with social goals

The principle of rationality states that agents will take efficient means to achieve their goals, given their beliefs about the world. We assume that agents receive rewards that depend on their goals, their own states and the states of their conspecifics, and that agents plan to maximize their expected

future rewards minus the costs they incur, potentially using a predictive model of the other agent's behavior.

Let $\mathcal{R}(W, G_i)$ be an agent's real-valued reward function, defined over the state of the world and the agent's social goal. We assume the reward function for each goal is given by a linear function of the geodesic (shortest-path) distance $\text{geo_dist}(i, j)$ between agent i and agent j :

$$\mathcal{R}(W, \text{Approach}) = -\alpha \cdot \text{geo_dist}(i, j), \quad (4.3)$$

$$\mathcal{R}(W, \text{Avoid}) = \alpha \cdot \text{geo_dist}(i, j). \quad (4.4)$$

Let $\mathcal{C}(S_i^t, S_i^{t+1}, E)$ be the cost of moving from state S_i^t to S_i^{t+1} in environment E , assumed to be proportional to the Euclidean distance of the movement. The rational behavior for an agent is defined as the sequence of actions that maximizes the expected sum of future rewards, while minimizing the sum of incurred costs. This is given by the *value function*, defined over the complete state of the world, i 's goal, and i 's belief about j 's goal:

$$V(W, G_i, B_i(G_j)) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t \left(\mathcal{R}(W^t, G_i) - \mathcal{C}(S_i^{t-1}, S_i^t, E) \right) \middle| W^0 = W \right]. \quad (4.5)$$

The *discount rate* γ determines how much more the agent values immediate rewards than future rewards.

Computing the expectation involved in Equation 4.5 requires averaging over all future state sequences, conditioned on W_t , G_i , and i 's expectation of j 's behavior. Because L1 agents do not represent other agents' social goals (for L1 agents, $B_i(G_j) = \text{Null}$), their model of other agents is given by a uniform distribution over available actions, yielding a random walk through the state space. For L2 agents, representing other agents' goals allows anticipation of their likely future actions. We assume that L2 agents represent others as L1 agents, and first compute others' likely L1 plans, then plan their own actions contingent on other agents' predicted behavior. For both L1 and L2 agents, Equation 4.5 can be efficiently approximated using dynamic programming algorithms (Bertsekas, 2001).

To maximize their expected rewards, agents must choose valuable actions that incur low cost

and lead to states with high value. The value of an action is given by the expected discounted future rewards after action a_i^t , minus the expected cost of the action:

$$Q(W^t, a_i, G_i, B_i(G_j)) = \mathbb{E}_{P(W^{t+1} | W^t, a_i)} \left[\gamma \cdot V(W^{t+1}, G_i, B_i(G_j)) - \mathcal{C}(S_i^t, S_i^{t+1}, E) \right]. \quad (4.6)$$

We assume that agents maximize rewards probabilistically, and sometimes only approximate the optimal sequence of actions. The probability that agent i will take action a_i^t from world state W^t is:

$$P(a_i^t | G_i, B_i(G_j), W^t) \propto \exp(\beta \cdot Q(W^t, a_i^t, G_i, B_i(G_j))). \quad (4.7)$$

Given equation 4.7, Equation 4.1 is computed as the product of the probability of each agent's actions over all time steps.

4.2.2 Cue-based models

In addition to the theory-based models L1 and L2, we consider two cue-based models inspired by a “simple heuristics” account of human social goal inference (Blythe et al., 1999; Zacks, 2004). The computational models proposed by Blythe et al., 1999 and Zacks, 2004 both rely on features of agents’ movements that can be computed efficiently from local information, such as the gradient of the distance between agents. We assume a slightly more sophisticated approach in which agents follow the gradient of a linear function of the geodesic distance to the other agent, minus a “cost” given by the step size.

We denote our cue-based alternatives H1 and H2. We assume that H1 agents do not take the environment E into account when computing the geodesic distance to the other agent. H2 does consider the environmental constraints when computing geodesic distance, and offers a quick-and-dirty alternative to a planning-based approach to incorporating the environment into predicting goal-directed behavior.

Let $\mathcal{R}(W^t, G_i)$ and $\mathcal{C}(S_i^{t-1}, S_i^t, E)$ be defined as above. Then the probability of action a_i , given

the world state and goal G_i for an H1- or H2-agent is:

$$P(a_i^t | G_i, B_i(G_j), W^t) \propto \exp(\beta \cdot \mathbb{E}_{P(W^{t+1} | W^t, a_i^t)} [\mathcal{R}(W^{t+1}, G_i) - \mathcal{C}(S_i^t, S_i^{t+1}, E)]). \quad (4.8)$$

This model may seem similar to our theory-based models, but there is a crucial difference: it does not require a rational planning-based model of agents' decision making process, but instead regards actions as a function of the goal and the environment alone.

4.3 Experiment

Our experiment collected human judgments in a task of categorizing agents as “approaching” or “avoiding” one another given short observations of their interactions in a simple, two-dimensional maze-like environment. The experiment was designed to distinguish the predictions of cue-based models H1 and H2 and theory-based models L1 and L2 by simultaneously varying local motion cues and the global environmental context to determine whether either class of models could account for the dependence of people’s judgments on these factors. We hypothesized that only the theory-based account would be able to explain the interaction between people’s sensitivity to subtle environmental changes and the highly salient cues provided by agents’ action sequences.

4.3.1 Method

Participants

Participants were 20 members of the MIT subject pool, 13 female and 7 male.

Stimuli

Our stimuli were designed following previous research showing that simple two-dimensional animations of moving shapes evoke strong impressions of animacy and inferences of mental states (Tremoulet & Feldman, 2000). Each stimulus displayed two agents, represented by red and green dots respectively. As agents moved along their trajectories, smaller dots of the same color trailed behind

them, recording their path history. Agents' paths were either 4 steps or 6 steps long. The environment was a discrete grid of 23 squares wide by 23 squares high, with walls represented as solid black lines. Agents' movements were restricted to adjacent squares, with directions {N,S,E,W,NE,NW,SE,SW}. All experimental stimuli are shown in Fig. 4-2.

Our stimulus design combined four different conditions of agents' movement with 7 different environmental and social contexts. The four joint movement conditions were: both agents moving toward each other ($\rightarrow\leftarrow$), both agents moving to the right ($\rightarrow\rightarrow$), both agents moving to the left ($\leftarrow\leftarrow$), and both agents moving away from each other ($\leftarrow\rightarrow$). These movement conditions are shown across the rows of Fig. 4-2.

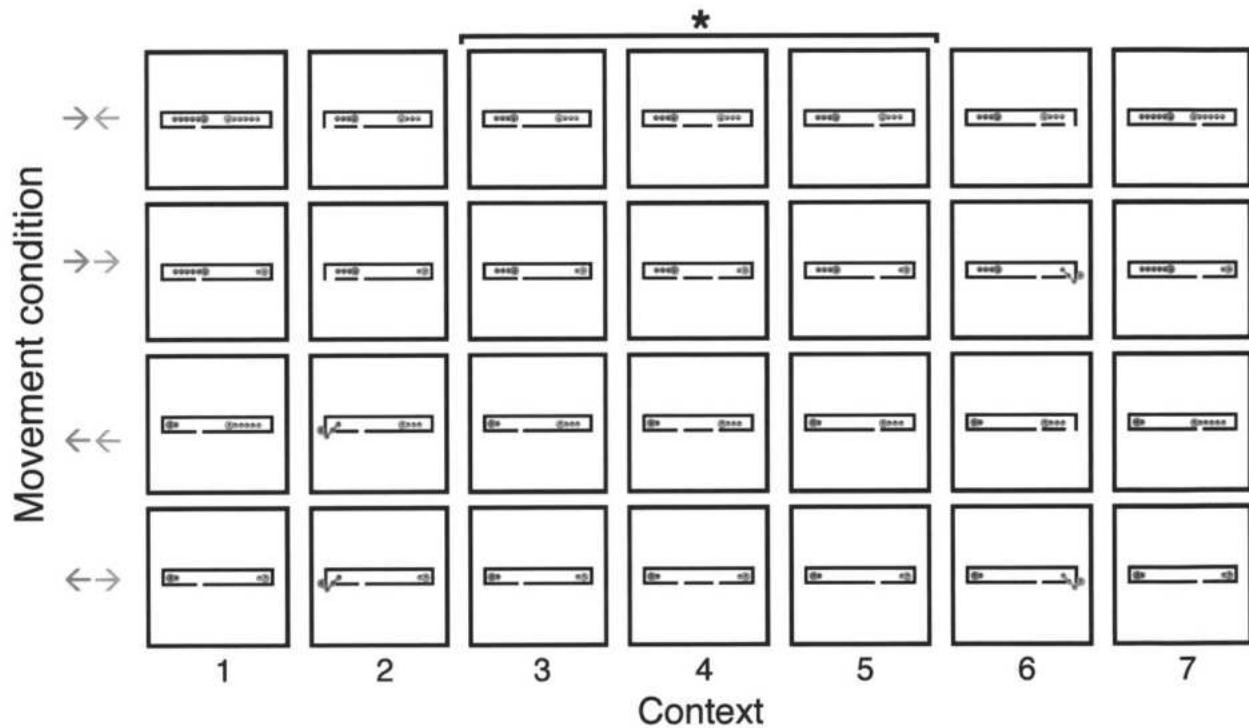


Figure 4-2: Experiment 1 stimuli.

The 7 different contextual conditions are shown across the columns of Fig. 4-2. Conditions 2-6 vary the environmental context by modifying the pattern of gaps in the wall of the maze. Conditions 1 and 7 modify the length of the agents' paths, displaying two more steps than the other conditions. Up-down orientation of the display was counterbalanced within subjects, yielding 56 stimuli in total. The side on which red and green agents appeared, and the order of the

“Avoiding/Aproaching” options were randomized between subjects (but constant throughout the experiment for individual subjects). Stimuli were presented in random order.

Procedure

Subjects were first presented with a cover story about intelligent aliens interacting in their natural environment. They were told that the aliens could not move through walls but that they could move through gaps in walls, and that they could see past walls and knew the complete state of the world at all times. Lastly, they were told that they would be watching videos of the aliens' movements and that they would rate whether each alien was trying to approach the other trying to avoid the other, or whether they couldn't tell. During the experiment, subjects viewed short animations of agents interactions and made ternary forced-choice decisions between the options “Approaching”, “Avoiding” and “Can’t tell” for both agents in each stimulus.

Modeling

We applied Equation 4.2 to model people’s judgments in our experimental task. However, for L2, because this equation specifies a joint probability distribution over both agents’ goals and their beliefs about each other’s goals, we marginalized over each L2 agent’s beliefs about the other agent’s goal.

For each stimulus, the model provided the probability that each agent had goal **Approaching** (the probability that each agent had goal **Avoiding** was just the complement of this). To compare people’s judgments to these model predictions, we first coded the **Approaching** rating as 1, the **Avoiding** rating as 0, and the “Can’t tell” rating as 0.5. We then averaged over the counterbalanced orientation conditions within subjects, and averaged over ratings between subjects and computed the standard error values of these averages.

4.3.2 Results

Results of the experiment are shown in Fig. 4-3. People’s judgments are shown in column 1, and demonstrate strong sensitivity to environmental factors. For instance, in condition $\rightarrow\leftarrow$, in contexts 3 and 4 people rated the Red agent’s movements as ambiguous, despite the fact that it was moving directly toward the Green agent in these conditions. Likewise for contexts 4 and 5 for the Green agent. Column 2 shows that the theory-based model L2 predicted this inference, but in column 3, the cue-based model did not.

In addition to predicting the overall trend of people’s inferences in condition $\rightarrow\leftarrow$, the L2 model also provides a plausible explanation for why people made these inferences. For instance, in context 3, Red’s action appears to be an equally sensible option when it is chasing or fleeing, which the model confirms. In contrast, because H2 does not incorporate planning into its predictions, it assumes that agents with the goal to avoid will head directly away from the other agent, and that agents with the approach goal will head directly for the other agent. Because of this, H2 makes the same predictions in every context for the $\rightarrow\leftarrow$ condition, and indeed, follows this pattern for the other movement conditions as well.

4.4 Conclusion

How can we reason about the social goals of other people, effortlessly inferring that “she’s chasing him” and “he wants her to notice him”? On one side, bottom-up approaches focus on simple, easily computed cues such as the relative motion between two agents. On the other side, theory-based approaches cast these inferences as top-down processes drawing on abstract knowledge and sensitive to context and background knowledge. The theory-based approach is attractive for its promise of capturing more realistic interactions, but can be quite difficult to interpret in a precise, computational way.

This paper presented a family of theory-based models of social goal inference. These models all embody the principle of rationality and use inverse planning as their core inferential engine, but vary in the sophistication of representations required. At one end, the simplest of these models

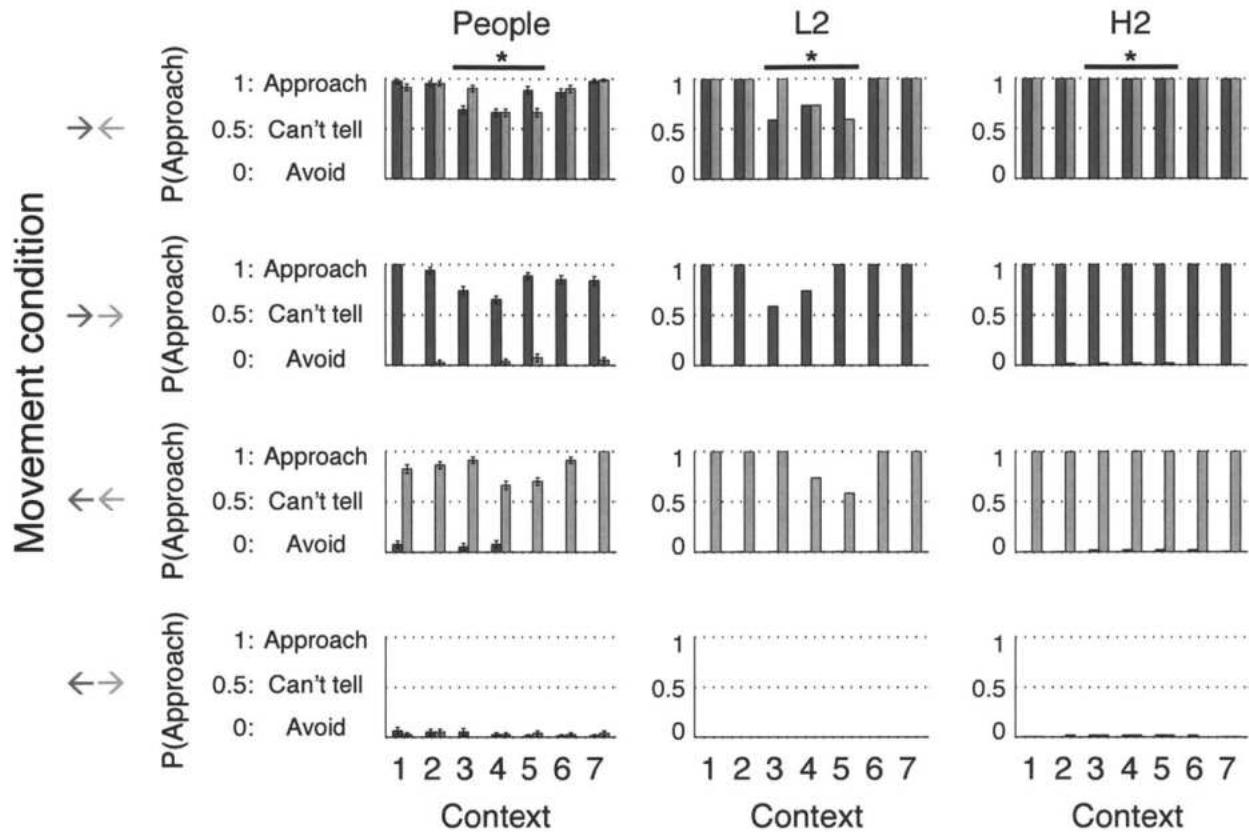


Figure 4-3: Experiment 1 results. Average subject ratings and model predictions. The labels in this figure directly correspond to the stimulus numbers in Fig. 4-2.

allows agents to represent only the properties of other agents (such as location), but not their goals. The most sophisticated of these models realizes second-order social goal inference: reasoning about agents' representations of other agents' goals. We distinguished simple cue-based models from our theory-based models using an experiment based on simple motion patterns of pairs of agents. We further argued that the simplest level of representation is inadequate to account for human social goal inferences—second-order content is necessary to capture such thoughts as “he wants me to believe him”.

Further work is needed to better distinguish between first- and second-order content in social goal inference, and to understand the factors that affect people’s use of these various representational abilities. More generally, the inverse planning approach can be applied to a much wider range of social goals than we have considered here. We are particularly interested to model goals

such as “helping” or “hindering”, in which agents take as their reward function some function of another agent’s reward. Inference to these goals is more computationally challenging but still falls under the scope of our inverse planning framework, and has also recently been shown to be within the reach of human infants.

Chapter 5

Conclusion

This thesis gave a computational account of aspects of the form, content, and function of human Theory of Mind (ToM). It was argued that at the heart of ToM is a family of probabilistic, causal models of intentional agency, which support joint behavior prediction and mental state inference. Intuitive principles govern the structure and content of these models in a context-sensitive manner, generating theories of intentional agents' mental states, their relation to the subjective situation, and their effect on behavior. Specifically, these principles were provided by versions of formalisms for belief-, utility-, and goal-dependent planning or decision-making, originally developed as normative models of rationality by economists and computer scientists, and adapted here to capture people's conception of intentional agency.

In an array of experiments, versions of the Bayesian ToM (BToM) framework provided precise, accurate predictions of human judgments, at both the group and individual levels. In each case, the relative performance of alternative models, each inspired by prominent proposals from the social perception literature, was assessed. Overall, the alternative models performed poorly compared with BToM, demonstrating the difficulty of capturing human social judgments, and the success of BToM in meeting this challenge.

These results suggest that BToM provides the "right" or at least "best" account of human social reasoning within the present experiments, but in what precise sense could this be (or not be) the case in general? From the perspective of ideal-observer analysis (Geisler, 1989; Knill & Richards,

1996) or rational analysis (Anderson, 1990; Chater & Oaksford, 1999), the question hinges on whether economic models of expected utility-maximization are actually the correct models of naturalistic behavior within the current settings. If the probabilistic generative models posited by BToM match the statistics of the natural phenomena, then Bayesian inference is the optimal solution to the problem the observer faces, and human subjects perform impressively close to the ideal limit in the experimental settings presented in this thesis.

But are economic models of rational decision-making – e.g., MDPs, POMDPs and Markov games – really the correct generative models for behavior? Does BToM really provide an ideal-observer analysis? In the present experimental settings, perhaps. Goal-directed navigation (Chapter 2), foraging (Chapter 3), and pursuit/evasion (Chapter 4) are very natural tasks, even for infants. They are evolutionarily ancient and necessary for the survival of our species (Cosmides & Tooby, 1992; Bogdan, 1997; Barrett et al., 2005; Krasnow et al., 2011), and we gain volumes of experience performing and observing others perform them throughout the lifespan. Therefore, it is plausible that humans both behave near-optimally, and have finely-tuned expectations that others will behave near-optimally in these key contexts. This thesis has provided evidence for the latter claim, but it will be important to test the former empirically.

Although this thesis has argued for a particular form for the forward model of ToM in terms of approximately rational planning, two other potential mechanisms are consistent with the results presented here: statistical estimation of behavior patterns, and simulation. Like planning-based models, these alternatives are consistent with an accordance between actual and expected behavior within naturalistic contexts, and make similar inferences in the present experiments. The first alternative, based on statistical estimation of the mental-state contingent distribution of behavior directly from data, will be most effective in domains where behavioral observations are readily available. This approach could achieve limited generalization to novel domains via state-abstraction – clustering similar states together based on their perceptual features (or their affordances) to extend behaviors observed in one state to predict behavior in unobserved states. On this account, people’s expectations about the experimental contexts presented in this thesis derive from their extensive experience observing intentional action in spatial contexts, which is extended to 2D animated sce-

narios using spatial similarity. A particular challenge for the statistical account is explaining how a volume of information about others' mental states and actions sufficient to reliably estimate these patterns is obtained. For simple, object-directed goals, there is often a directly observable association between action and outcome, perhaps providing a sufficient cue for learning. For complex arrangements of beliefs, desires and intentions, joint estimation of statistical dependencies is a hard inference problem, requiring substantial computational power and access to vast stores of data.

The second alternative relates to the simulation theory (A. I. Goldman, 2006) of mental state attribution. A version of simulation theory consistent with approximately Bayesian reasoning could posit that the predictive model of others' actions results from simulating one's own decision-making within an agent's perceived situational context. Inference could then be performed using a "generate and test" procedure – generating hypothetical mental states, simulating the resulting action, and testing against the actual observed actions of the agent, with mental states which best match the observed actions considered the most likely.

Under the proposition that humans behave near-optimally in the contexts considered in this thesis, both alternative mechanisms – statistical estimation and simulation – will produce similar action predictions in the forward direction and mental state inferences in the inverse direction. This suggests a way to distinguish the three possibilities: identifying contexts in which there is a dissociation between people's expectations of behavior and the behavior they produce. Indeed, this analysis was proposed by Saxe (2005), whose "argument from error" (against simulation theory) cites a body of evidence showing that human action predictions often diverge from well-known behavioral phenomena. In these contexts, people's expectations exhibit systematic "over-attribution" of rationality, supporting the role of a rationality-based intuitive theory, rather than a simulation-based mechanism, which is inconsistent with such differences between actions and expectations. A similar argument can be mounted against the statistical estimation alternative; its errors will tend toward weak, uncertain predictions in unfamiliar domains.

Although over-attribution of rationality constitutes a departure from ideal-observer or rational inference in the sense defined above, an advantage of predictive models of rational planning is their ability to generalize to novel contexts, allowing behavior to be understood far outside of the

familiar situations encountered in everyday life. Furthermore, the rationality assumption may provide a general space of prior models which is refined as more data from others' or even one's own behavior are observed. Additional data could even be "simulated" offline using neural mechanisms for decision-making; these possibilities highlight the prospect of a hybrid account combining the strengths of planning-, statistics- and simulation-based mechanisms within the BToM framework. Elucidating the role of these and other mechanisms within a range of contexts by comparing actual behavior, predicted behavior, and interpreted behavior is a promising next step within the BToM research program defined in this thesis.

Appendix A

Formal details of goal inference modeling

A.1 Markov decision problems

This section formalizes the encoding of an agent’s environment and goal into a Markov decision problem (MDP), and describes how this MDP can be solved efficiently by algorithms for rational planning. Let π be an agent’s plan, referred to here (and in the MDP literature) as a *policy*, such that $P_\pi(a_t|s_t, g, w)$ is a probability distribution over actions a_t at time t , given the agent’s state s_t at time t , the agent’s goal g and world state w . This distribution formalizes $P(\text{Actions}|\text{Goal}, \text{Environment})$, the expression for probabilistic planning sketched in the main text. The policy π encodes all goal-dependent plans the agent could make in a given environment.

We assume that agents’ policies follow the principle of rationality. Within a goal-based MDP, this means that agents choose action sequences that minimize the expected cost to achieve their goals, given their beliefs about the environment. Let $C_{g,w}(a, s)$ be the environment- and goal-dependent cost to an agent of taking action a in state s . The expected cost to an agent of executing policy π starting from state s is given by the agent’s *value function*, which sums the costs the agent is expected to incur over an infinite horizon:

$$V_{g,w}^\pi(s) = E_\pi \left[\sum_{t=1}^{\infty} \sum_{a_t} P_\pi(a_t|s_t, g, w) C_{g,w}(a_t, s_t) \middle| s_1 = s \right]. \quad (\text{A.1})$$

In general, cost functions may differ between agents and environments. For the environments we consider, action costs are assumed to be proportional to the negative length of the resulting movement, and the Stay action incurs a small cost as well. We assume that agents stop incurring costs once they reach their goals, implying that rational agents will try to reach their goals as quickly as possible.

The state-action value function, or Q , defines the expected cost of taking action a_t from state s_t and executing policy π afterwards by averaging possible outcomes s_{t+1} caused by a_t :

$$Q_{g,w}^\pi(s_t, a_t) = \sum_{s_{t+1}} P(s_{t+1}|s_t, a_t, w) V_{g,w}^\pi(s_{t+1}) + C_{g,w}(a_t, s_t). \quad (\text{A.2})$$

Here, $P(s_{t+1}|s_t, a_t, w)$ is the state transition distribution, which specifies the probability of moving to state s_{t+1} from state s_t , as a result of action a_t , in world w . For simplicity, we assume state transitions are deterministic in our experiments.

The optimal action from any state is computed by greedily choosing the action that maximizes the Q -function. However, instead of this deterministic policy, we assume that agents have a probability distribution over actions associated with policy π :

$$P_\pi(a_t|s_t, g, w) \propto \exp(\beta Q_{g,w}^\pi(s_t, a_t)). \quad (\text{A.3})$$

This type of policy is called a *Boltzmann policy*, which takes the “soft-max” of the Q -function, yielding an approximate principle of rationality, where β controls the amount of noise in the agent’s actions.

The optimal Boltzmann policy π^* and the value function of π^* satisfy the Bellman equation (Bellman, 1957) for all states:

$$V_{g,w}^{\pi^*}(s) = \sum_{a_t} P_{\pi^*}(a_t|s_t, g, w) Q_{g,w}^{\pi^*}(s_t, a_t). \quad (\text{A.4})$$

These equations can be solved efficiently using the value iteration algorithm (Bertsekas, 2001), which iteratively updates the left- and right-hand sides of Equation A.4 for all states until conver-

gence to a unique fixed point.

Given an agent’s situation-specific policy, goal inferences and goal-based action predictions depend crucially on the prior over goals, corresponding to $P(\text{Goal}|\text{Environment})$ in the main text. This prior is instantiated by different models within the inverse planning framework. In the next section, we describe three inverse planning models based on different hypotheses about people’s prior knowledge about goals, denoted $M1(\beta)$, $M2(\beta, \gamma)$, and $M3(\beta, \kappa)$, which roughly correspond to the three kinds of explanations we offered for the woman’s anomalous behavior in the introductory vignette in the main text. In addition to these models, we also consider a simple heuristic alternative, denoted $H(\beta)$, that people might apply in action understanding. For each model, we will describe the computations involved in each of our three key tasks: online goal inference, retrospective goal inference and goal-based action prediction.

These three kinds of inferences all depend on applying Bayes’ rule, in which we compare all possible goal hypotheses against each other, in terms of how well they explain an observed action sequence. Each goal hypothesis yields a different MDP that must be solved. In this paper, we exhaustively enumerate the full hypothesis space of goals under each model, solving the corresponding MDP for each one. This is computationally feasible for the simple environments and the restricted hypothesis spaces of goals that we consider here. Real world settings will likely require more complex goal priors which allow goals to be selected from a very large or infinite hypothesis space. In this case, full enumeration will not be possible, and efficient and accurate approximate inference schemes will be required (Ramachandran & Amir, 2007; Todorov, Under Review). Likewise, in MDPs defined over complex environments with many state variables, the state space quickly grows too large to solve the MDP exactly. In these cases, approximate rational planning algorithms are necessary, and this is an active area of research (Ng & Jordan, 2000; Guestrin et al., 2003; Wingate & Seppi, 2005; Todorov, 2006) within the field of artificial intelligence.

A.2 Inverse planning models

A.2.1 Model 1: single underlying goal

Our first model, denoted $M1(\beta)$, is our most basic instantiation of inverse planning. $M1$ assumes that agents have one underlying goal in each action sequence that must be inferred. A graphical model of $M1$ is shown in Fig. A-1(a). Observations of agents are assumed to begin at time $t = 1$, with the agent occupying state s_1 in environment w . The agent is assumed to have an invariant goal g , which generates actions a_1 through a_{T-1} according to the policy from Equation A.3, where the parameter β determines the agent's level of determinism. At high values of β , agents rarely deviate from the optimal path to their goals, but at low β values, agents' behavior is noisy, becoming a random walk at $\beta = 0$. Agents' actions generate state transitions, producing the state sequence s_1, s_2, \dots, s_T . The objective of inverse planning is to invert this model, yielding inferences of agents' goals g , given observations of the state sequence $s_{1:T}$.

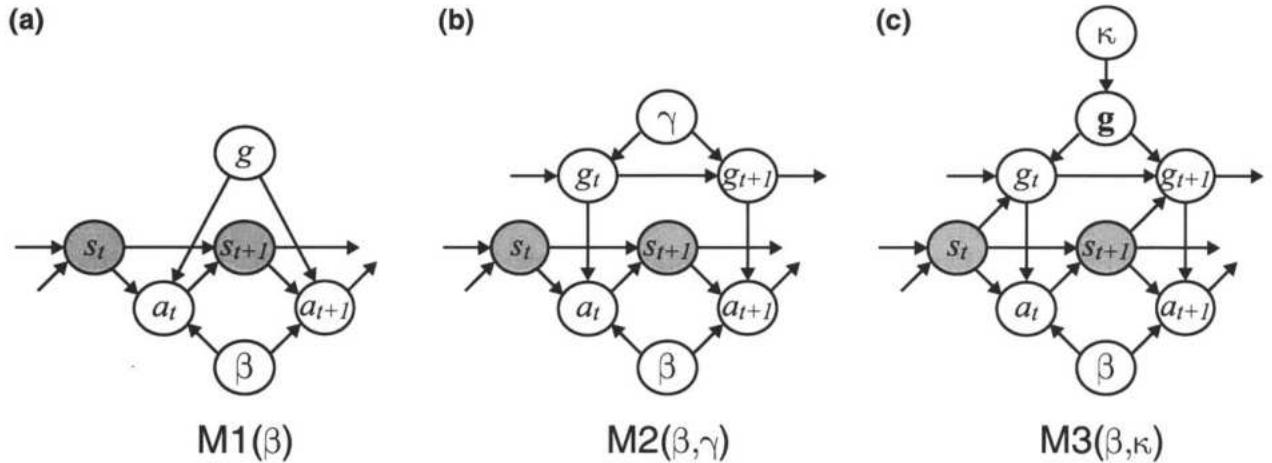


Figure A-1: Graphical models of $M1$, $M2$ and $M3$. All variables except β , γ and κ are implicitly conditioned on the environment w .

Given an observed state sequence and the environment, the distribution over the agent's goal in $M1$ is computed using Bayes' rule:

$$P(g|s_{1:T}, w) \propto P(s_{2:T}|s_1, g, w)P(g|w), \quad (\text{A.5})$$

where $P(g|w)$ is the prior over the agent's goal. There are many possible ways to treat $P(g|w)$, but in our experiments, for simplicity, we will assume it is uniform over the set of goals that are realizable in the environment w , *i.e.* each position in the grid not blocked by an obstacle. In Equation A.5, the probability of a state sequence given a goal and the environment, $P(s_{2:T}|s_1, g, w)$, is computed as the product of the probabilities of the individual state transitions from s_t to s_{t+1} , given the goal and the environment:

$$P(s_{2:T}|s_1, g, w) = \prod_{t=1}^{T-1} P(s_{t+1}|s_t, g, w). \quad (\text{A.6})$$

In Equation A.6, we marginalize over actions to compute the probability of the next state s_{t+1} , given the current state s_t , the goal g , and the environment w :

$$P(s_{t+1}|s_t, g, w) = \sum_{a_t \in A_{s_t}} P(s_{t+1}|s_t, a_t, w) P_\pi(a_t|s_t, g, w). \quad (\text{A.7})$$

Equation A.5 corresponds directly to Equation 1 in the main text, and allows us to model people's online and retrospective goal inferences in Experiments 1 and 2.

Prediction of future actions, given observations of previous actions, is done by computing the probability of a novel state sequence $s'_{1:T}$ in the environment w , conditioned on a set of N observed state sequences of length T $s_{1:T}^{1:N}$ in w . Assuming the agent's goal is unchanged, this is just an average over the predicted actions for each possible goal g , weighted by the posterior probability of g given the observed sequences:

$$P(s'_{2:T}|s'_1, s_{1:T}^{1:N}, w) = \sum_g P(s'_{2:T}|s'_1, g, w) P(g|s_{1:T}^{1:N}, w). \quad (\text{A.8})$$

Equation A.8 corresponds directly to Equation 2 in the main text, and allows us to model people's action predictions in Experiment 3.

Unlike the more complex models presented below, M1 explains agents' actions in terms of a single underlying goal, and accounts for all unexplained deviations from the shortest path to the goal in terms of random departures from optimality. However, there is often not a single invariant

goal that fully explains an agent’s actions. Instead, real-world agents typically pursue goals with rich structure, such as subgoals, multiple objectives, and goal changes, as in the introductory example. Next, we describe models M2 and M3, which assume more complex goal structures. These models generalize M1: M1(β) is a special case of both M2(β, γ) and M3(β, κ), with γ or κ equal to 0.

A.2.2 Model 2: changing goals

The next model we consider extends M1 by assuming that agents’ goals can change over time. We denote this model M2(β, γ), where γ is the probability that the agent will change its goal at each time. The graphical model of M2, shown in Fig. A-1(b), expands on the graphical model of M1 by replacing the single node g with a Markov chain representing the probability distribution over the sequence of goals g_1, g_2, \dots, g_{T-1} .

In M2, goals are indexed by time, and we modify Equation A.5 for M1 to represent $P(g_t|s_{1:T}, w)$, the posterior probability of an agent’s goals at each time, with $t < T$. In the online case, goal inferences are provided by the posterior distribution over goals at time t , given a state sequence $s_{1:t+1}$ and world w . To compute this, we recursively define the *forward distribution*:

$$P(g_t|s_{1:t+1}, w) \propto P(s_{t+1}|g_t, s_t, w) \sum_{g_{t-1}} P(g_t|g_{t-1}, w) P(g_{t-1}|s_{1:t}, w), \quad (\text{A.9})$$

where the recursion is initialized with $P(g_1|w)$, the probability distribution over initial goals at time $t = 1$, before any state transitions have been observed. $P(g_t|g_{t-1}, w)$ is the conditional distribution over changing to goal g_t from g_{t-1} . We assume that $P(g_1 = i|w) = \theta_i$, and that $P(g_t = i|g_{t-1} = j, w) = (1 - \gamma)\delta_{ij} + \gamma\theta_i$, where θ is typically assumed to be uniform over the set of goals¹.

Equation A.9 is an online version of Equation 1 in the main text, which allows us to model people’s online goal inferences in Experiment 1. When $\gamma = 0$, M2 reduces to M1, and goal

¹Note that this implies that given a goal change, the agent sometimes chooses the same goal as before, and only chooses a *new* goal with probability $\gamma(K - 1)/K$. We choose this parameterization for clarity and consistency, but M2 can easily be reparameterized by $\gamma' \leftarrow \gamma K/(K - 1)$.

changing is prohibited. When $\gamma = 1$, the model is equivalent to randomly choosing new goal i with probability θ_i at each time step. Intermediate values of γ between 0 and 1 interpolate between these extremes.

The retrospective version of Equation A.5 for M2 is given by $P(g_t|s_{1:T}, w)$, the marginal probability of a goal at time t given the entire state sequence $s_{1:T}$, with $t < T - 1$. To compute this, we use a variant of the forward-backward algorithm (Rabiner, 1989). The forward distribution is defined by Equation A.9 above. The *backward distribution* is recursively defined by:

$$P(s_{t+2:T}|g_t, s_{1:t+1}, w) = \sum_{g_{t+1}} P(g_{t+1}|g_t, w) P(s_{t+2}|g_{t+1}, s_{t+1}, w) P(s_{t+3:T}|g_{t+1}, s_{t+2}, w). \quad (\text{A.10})$$

The marginal probability of goal g_t given the state sequence $s_{1:T}$ is the product of the forward and backward distributions:

$$P(g_t|s_{1:T}, w) \propto P(g_t|s_{1:t+1}, w) P(s_{t+2:T}|g_t, s_{1:t+1}, w). \quad (\text{A.11})$$

Equation A.11 is a smoothed version of Equation 1 in the main text, allowing us to model subjects' retrospective goal inferences in Experiment 2. The parameter γ plays a key role in retrospective inferences, determining how information from past and future movements is integrated into the distribution over current goals. When $\gamma = 0$, M2 reduces to M1: changing goals is prohibited, and future information constrains the probability of all past goals to be equal to $P(g_{T-1}|s_{1:T}, w)$. When $\gamma = 1$, only the movement from s_t to s_{t+1} carries information about g_t ; all other past and future movements carry no information about the current goal.

Prediction of agents' future actions in M2 differs from prediction in M1 because agents' goals are no longer assumed constant across observation sequences. Rather than inferring an invariant goal, in M2, prediction of agents' future actions, given their past actions, is done by inferring θ . The posterior distribution over θ , given a set of N observed state sequences of length T $s_{1:T}^{1:N}$ in environment w is computed by Bayes' rule:

$$P(\theta|s_{1:T}^{1:N}, w) \propto P(s_{2:T}^{1:N}|s_1^{1:N}, \theta, w) P(\theta|w). \quad (\text{A.12})$$

The probability of a novel state sequence $s'_{1:T}$, given environment w and previously observed state sequences $s^{1:N}_{1:T}$ in w is then:

$$P(s'_{1:T}|s^{1:N}_{1:T}, w) = \int_{\theta} P(s'_{2:T}|s'_1, w, \theta) P(\theta|s^{1:N}_{1:T}, w) d\theta. \quad (\text{A.13})$$

Equation A.13 corresponds to Equation 2 in the main text. Because the integral in Equation A.13 is intractable, we use a grid approximation to integrate over θ to model people's judgments in Experiment 3².

A.2.3 Model 3: complex goals

Our final model extends M1 by assuming that agents' goals can be *complex*, and can include the constraint that agents must pass through a sequence of "subgoals" along the way to their end goals. We denote this model $M3(\beta, \kappa)$. A graphical model of M3 is shown in Fig. A-1(c).

In M3, complex goals g are a list of states, instead of a single state g , as in M1. Let g be a complex goal with 0 or more subgoals, represented as a list of length $\dim(g)$, where $\dim(g) \geq 1$. $P(g|w)$ is the prior over complex goals. The prior assumes that with probability κ , agents pick a subgoal uniformly from the state space, and with probability $1 - \kappa$, agents do not choose a subgoal. Agents continue choosing additional subgoals with probability κ until the sampling terminates with probability $1 - \kappa$, implying that number of subgoals chosen is geometrically distributed. Then, agents pick an end goal uniformly from the state space. M3 includes M1 as a special case with $\kappa = 0$, implying that the agent picks 0 subgoals.

The posterior probability of a complex goal under this model is computed identically to M1, except now g is a complex goal:

$$P(g|s_{1:T}, w) \propto P(s_{2:T}|s_1, g, w) P(g|w). \quad (\text{A.14})$$

The probability of an agent's state sequence, given a complex goal g and environment w , is com-

²Another possible approximation is to compute the maximum likelihood (ML) estimate θ_{ML} , conditioned on state sequences $s^{1:N}_{1:T}$ and environment w using the EM algorithm, and use this ML estimate to predict hypothetical actions.

puted by segmenting the full state sequence into sequences in which the agent pursues particular subgoals. Once a subgoal is reached, a new segment begins in which the agent pursues its next subgoal. Formally,

$$P(s_{2:T}|s_1, \mathbf{g}, w) = \prod_{i=1}^{\dim(\mathbf{g})} \prod_{t=k_{i-1}}^{k_i} P(s_{t+1}|s_t, g_t = i, w), \quad (\text{A.15})$$

where $k_0 = 1$ and $k_i = \min(\{t|s_t = g_i \wedge t > k_{i-1}\})$, $i > 0$.

Inferences about an agent's end goal are obtained by marginalizing over goal types, and within the complex goal type, marginalizing over possible subgoals. Let $\mathbf{g} = [g_{1:\dim(\mathbf{g})-1} \ g_{\dim(\mathbf{g})}]$. Then

$$P(g_{\dim(\mathbf{g})}|s_{1:T}, w) \propto \sum_{g_{1:\dim(\mathbf{g})-1}} P(s_{2:T}|s_1, \mathbf{g}, w)P(\mathbf{g}|w) \quad (\text{A.16})$$

is the marginal probability of an end goal, given a state sequence and the environment. Equation A.16 corresponds to Equation 1 in the main text, and allows us to model people's online and retrospective inferences in Experiments 1 and 2.

Prediction in M3 is similar to prediction in M1, but instead of averaging over simple goals, we average over complex goals:

$$P(s'_{1:T}|s_{1:T}^{1:N}, w) = \sum_{\mathbf{g}} P(s'_{2:T}|s'_1, \mathbf{g}, w)P(\mathbf{g}|s_{1:T}^{1:N}, w). \quad (\text{A.17})$$

Equation A.17 corresponds directly to Equation 2 in the main text, and allows us to model people's action predictions in Experiment 3.

A.2.4 Heuristic alternative H

The heuristic alternative we consider, denoted $H(\beta)$, assumes that the probability of a goal is a function of only the agent's last observed movement. H is inspired by heuristics that categorize intentional movement based on temporally local motion cues, such as the rate of change of the relative distance between two objects. For the sake of comparison with our other models, we

formulate H as a limiting case of inverse planning. H is equivalent to a special case of M2, with probability of changing goals at each timestep $\gamma = 1$, implying that the agent chooses a new goal at each timestep. This yields a version of Equation A.11 that only depends on the agent’s last movement, regardless of the length of the agent’s movement history:

$$P(g_t|s_{1:t+1}, w) \propto P(s_{t+1}|s_t, g_t, w)P(g_t|w). \quad (\text{A.18})$$

In contrast, Equation A.5 for M1, Equation A.11 for M2, and Equation A.16 for M3 infer a goal based on how well it explains the agent’s entire state sequence, rather than just the last movement.

A.3 Experiment 1

A.3.1 Bootstrap cross-validated correlational analyses

We assessed the statistical significance of the differences between M1, M2, M3 and H with a bootstrap cross-validated (Cohen, 1995) correlational analysis. Bootstrap cross-validation (BSCV) is a non-parametric technique for model selection, which measures the goodness-of-fit of models to data while preventing overfitting and controlling for model complexity. For a large number of iterations N , our analysis selected random training subsets of size k of participants’ data (by sampling uniformly with replacement from all data points) and found the parameters of each model that yielded the highest correlations with these training datasets. For each iteration, we then computed the correlation of the model fit on that iteration with a testing dataset given by the complement of the training dataset on that iteration. Across all iterations, we computed the average correlation $\langle r \rangle$ of each model with the testing datasets, and we computed the proportion of iterations in which the correlation of one model was greater than another model. This allowed us to estimate the goodness-of-fit of each model class, and to compute p -values for significance tests of whether one model class predicted people’s data better than another. Our BSCV analyses of Experiment 1 (overall and targeted) used parameters $N = 10000$ and $k = 50$.

A.3.2 Analysis of parameters

This analysis examined a grid of parameter values for each model. We tested each model using 10 evenly spaced β values from 0.5 to 5. For M2, for each value of β we tested 20 evenly spaced γ values between 0.05 and 1. For M3, for each value of β we tested 20 evenly spaced κ values between 0.05 and 1. For each model, a grid of correlation coefficients with people's judgments using these parameters is shown in Fig. A-2.

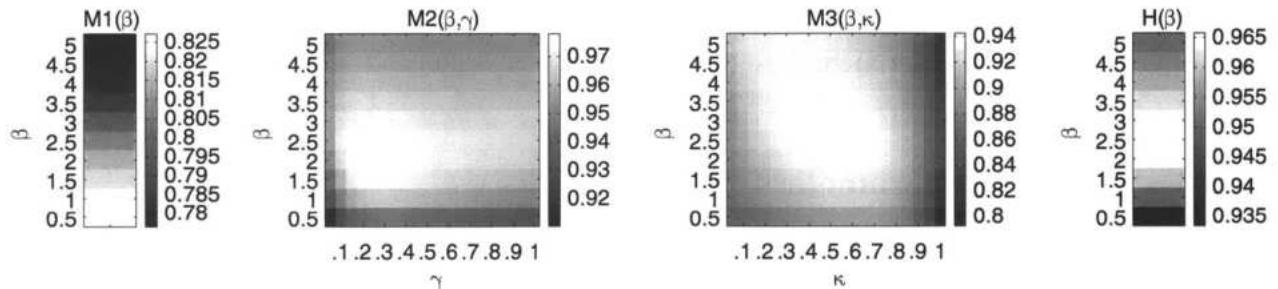


Figure A-2: Correlation grids for M1, M2, M3 and H for Experiment 1. These plots show the correlation of each model with people's data for a grid of parameter values. M2, M3 and H all correlated highly with people's ratings, with maximum correlations with people's data of $r > 0.97$, $r > 0.94$ and $r > 0.96$, respectively, while M1 correlated less highly with people's judgments, with a maximum correlation of $r > 0.82$.

For each model, a robust range of parameters yielded correlations that were close to the optimum. For M1, all parameter values yielded correlations that were low relative to M2, M3 and H. Among the parameter settings we tested for M1, low β values yielded the highest correlations with people's data, because M1 could only explain the indirect paths in our stimuli in terms of noisy actions.

For M2, all parameter values that we tested yielded correlations above $r = 0.9$. The optimal range for β was between 1.5 and 2.5, and the optimal range for γ was between 0.10 and 0.25. The worst correlations were obtained at β values below 1.0, which yielded a uniform distribution over actions in the limit at $\beta = 0$, and at γ values below 0.5, which yielded M1 in the limit at $\gamma = 0$. The other limit, at $\gamma = 1$, yielded H, for which the same range of β values yielded good performance as for M2.

For M3, changing the parameter values resulted in a wider numerical range of correlations

with subjects' data, from a minimum around $r = 0.8$, to a maximum near $r = 0.94$. M3 correlated relatively highly with people's judgments at values of β above 1.0 and values of κ below 0.8. The optimal range for β was between 1.0 and 3.0, and the optimal range for κ was between 0.2 and 0.7.

A.4 Experiment 2

A.4.1 Bootstrap cross-validated correlational analysis

Our BSCV analysis of Experiment 2 used parameters $N = 10000$ and $k = 50$.

A.4.2 Analysis of parameters

This analysis examined a grid of parameter values for each model, using the same values as our analysis of the parameters for Experiment 1. A grid of the correlation coefficients of each model with people's judgments using these parameters is shown in Fig. A-3.

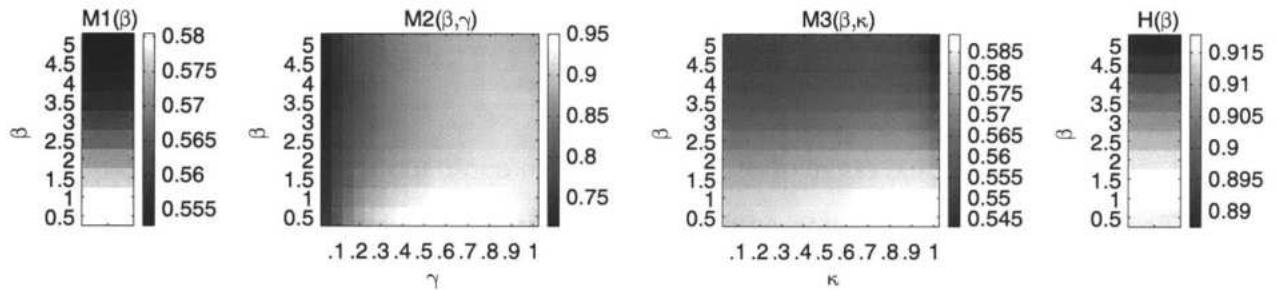


Figure A-3: Correlation grids for M1, M2, M3 and H for Experiment 2. These plots show the correlation of each model with people's data for a grid of parameter values. Models that could infer changing goals (M2 and H) correlated highly with people's ratings, with maximum correlations of $r > 0.95$ and $r > 0.91$, respectively. Models that could represent only static goals (M1 and M3) correlated less highly with people's judgments, both with maximum correlations with people's data of $r > 0.58$.

In Experiment 2, models based on changing goals (M2 and H) correlated highly with people's data, while models based on static goals (M1 and M3) correlated relatively poorly with people's data. For all the models we tested, β values below 1.0 yielded the best correlations with people's judgments, which was lower than in Experiment 1. For M2, γ values above 0.2 yielded the highest

correlations with people’s data, and the optimal range was between 0.4 and 0.7. For M3, the optimal range for κ was from 0.7 to 1.0.

A.5 Experiment 3

A.5.1 Data processing and model fitting

We modeled people’s judgments in Experiment 3 using the log posterior odds ratios between the two hypothetical actions in the response phase. Because the log posterior odds ratios varied between $-\infty$ and ∞ , and mean subject ratings varied between 1 and 9, we first mapped the log posterior odds ratios to the interval $(0, 1)$, then rescaled these values to the interval $(1, 9)$. Our mapping was based on the assumption that the log posterior odds ratios were normally distributed, and that subjects made their ratings by dividing the range of log posterior odds into 9 regions of equal probability, corresponding to the 9 points on the rating scale. To map the log posterior odds ratios to the interval $(0, 1)$, we first computed the z-scores of all log posterior odds ratios, then mapped these z-scores through the sigmoidal normal cumulative density function. We then computed correlation coefficients between these values and people’s ratings.

For M1 and M3, we assumed that the agent’s goal could be visible or invisible, and that the space of potential goals and subgoals was given by all grid squares in the environment. For M2, we restricted the space of goals to just the marked end goal and the potential subgoal of the particular condition. We modeled participants’ learning over training trials as inferring the probability of changing to either goal in M2. The probability of the agent choosing the end goal, given a goal change, was represented by the parameter θ , and the probability of choosing the subgoal was given by $1 - \theta$. To compute the integral over θ for posterior prediction of future actions, we approximated the continuous range of θ values from 0 to 1 using a discretization of 21 intervals.

A.5.2 Bootstrap cross-validated correlational analysis

Our BSCV analysis of Experiment 3 used parameters $N = 10000$ and $k = 20$.

A.5.3 Analysis of parameters

This analysis examined a grid of parameter values for each model, using the same values as our analyses of the parameters for Experiment 1 and Experiment 2. A grid of the correlation coefficients of each model with people’s judgments using these parameters is shown in Fig. A-4.

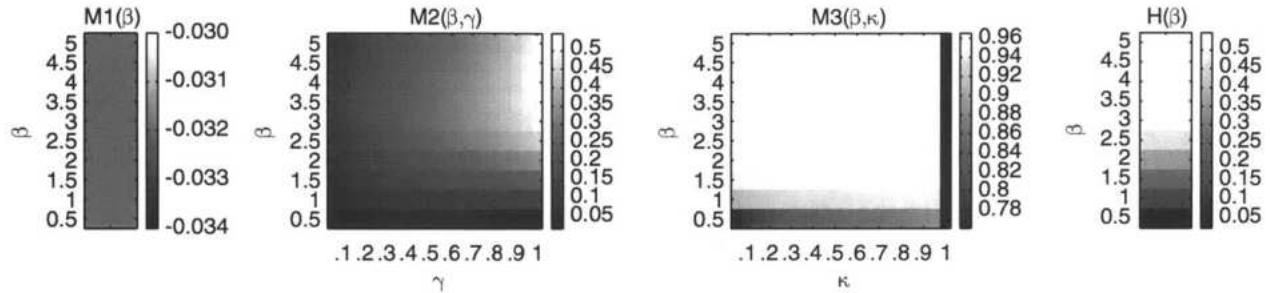


Figure A-4: Correlation grids for M1, M2, M3 and H for Experiment 3. These plots show the correlation of each model with people’s data for a grid of parameter values. Only M3 correlated highly with people’s judgments ($r > 0.96$), while M2 and H correlated relatively poorly with people’s data ($r > 0.5$), and the predictions of M1 were uncorrelated with people’s data ($r > -0.03$).

For Experiment 3, the predictions of M1 were uncorrelated with people’s judgments because M1 made the same prediction in each trial: that the agent would take the direct path to its goal. M1 made this prediction regardless of the value of β . For M2, M3 and H, higher β values yielded a better fit to people’s data than in the previous experiments, probably because the agents took straighter, less noisy paths in these stimuli.

For M2 and H, all parameter values yielded a low correlation coefficient with people’s judgments. Although our simulation of M2 (and H) used an approximation (described in Appendix A.2.2) this approximation faithfully captured the aspects of the stimuli that were most relevant for our subjects. Our approximation first restricted the hypothesis space of goals to include just the sub-goal and end goal from the stimuli, which was warranted by the fact that these goals explained the full range of agents’ actions in the stimuli. Second, our approximation of the integral from Equation A.13 was accurate, introducing a minimal amount of error into the model predictions.

For M2, high values of γ , which included H as a special case, performed best. High γ values allowed M2 to infer goal changes at every timestep, and to absorb the probability of switching

to the subgoal versus the end goal into the stimulus-dependent θ posterior. However, with this parameterization, M2 was still a poor predictor of people’s judgments. M2 could potentially predict a sequence of goals that first pursued the subgoal, and then pursued the end goal, but M2 lacked a strong inductive bias over the sequence of goals. Even if M2 learned from the training stimuli that the agent pursued the subgoal half of the time, and the end goal half of the time, it had no way to predict in what order the agent would pursue these goals in the test stimuli.

For M3, all values of κ from .05 to .95 performed well for β above 1.5. Values of β below 1.5 performed worst because they tended to explain indirect paths as random “noise”, rather than inferring a subgoal, and thus did not generalize evidence for subgoals from the training stimuli to the test stimuli. The correlation coefficient between M3 and people’s ratings was relatively insensitive to the value of κ because κ had an approximately additive effect on the log posterior odds ratio for moderate values, which did not affect the correlation.

A.6 General discussion

A.6.1 Hierarchical Bayesian analysis: model selection across experiments

Here we make precise the hierarchical Bayesian analysis laid out in the General Discussion of the main text for explaining how participants inferred which model or goal prior was appropriate to use in each experiment. Let D denote the ensemble of data that subjects saw during the course of a particular experiment, consisting of a pair of **Actions** and **Environment** for each of N trials, such that $D = \{\langle \text{Actions}_i, \text{Environment}_i \rangle | i \in 1, \dots, N\}$. The probability of a particular model M given D can then be computed using Bayes’ rule:

$$P(M|D) \propto P(D|M)P(M), \quad (\text{A.19})$$

where $P(M)$ is the prior probability of the model. The likelihood of the model $P(D|M)$ is the product of the marginal probability of the actions in each trial, conditioned on the environment in that trial and the model. To compute the marginal probability for each trial, we sum over all

possible goals because the agent's actions also depend on its goal:

$$P(D|M) = \prod_{i=1}^N \sum_{\text{Goals}} P(\text{Actions}_i|\text{Goal}, \text{Environment}_i, M) P(\text{Goal}|\text{Environment}_i, M).$$

The term inside the sum is just the probabilistic planning likelihood multiplied by the prior over goals, taken directly from Equation 1 in the main text. Thus, the likelihood of a model depends on how well it explains all the agents' actions across all trials.

To test whether people's inferences can be explained by rational inference in an HBM, we compared the log-marginal likelihood of M1, M2, M3 and H for each experiment, $\log P(D|M)$ from Equation A.19 (this is denoted $\log P(\text{Stimuli}|\text{Model})$ in the main text). This involves integrating over all degrees of freedom in the HBM, including all possible goals and the parameters β , γ and κ for each model, assuming a uniform prior.

References

- Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on machine learning*.
- Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, 82(2), 261–277.
- Albrecht, D. W., Zukerman, I., & Nicholson, A. E. (1998). Bayesian models for keyhole plan recognition in an adventure game. *User Modeling and User-Adapted Interaction*, 8(1-2), 5–47.
- Allen, J. F., & Perrault, C. R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15, 143–178.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, 20, 191–208.
- Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based social goal inference. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 1447–1455).
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113, 329–349.
- Baker, C. L., 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).
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal inference as inverse planning. In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (pp. 779–787).
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21, 37–46.
- 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.

sity Press.

- Bellman, R. (1957). *Dynamic programming*. Princeton, N.J.: Princeton University Press.
- Bello, P. (2011). Shared representations of belief and their effects on action selection: A preliminary computational cognitive model. In *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society* (pp. 2997–3002).
- Bello, P., Bignoli, P., & Cassimatis, N. L. (2007). Attention and association explain the emergence of reasoning about false beliefs in young children. In *Eighth international conference on cognitive modeling* (pp. 169–174).
- Bello, P., & Cassimatis, N. L. (2006a). Developmental accounts of theory-of-mind acquisition: Achieving clarity via computational cognitive modeling. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 1014–1019).
- Bello, P., & Cassimatis, N. L. (2006b). Some unifying principles for computational models of theory-of-mind. In *Seventh international conference on cognitive modeling*.
- Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, 4, 557–560.
- Bergen, L., Evans, O. R., & Tenenbaum, J. B. (2010). Learning structured preferences. In *Proceedings of the thirty-second annual conference of the cognitive science society*.
- Berthiaume, V. G., Onishi, K. H., & Shultz, T. R. (2008). A computational developmental model of the implicit false belief task. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 825–830).
- 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.
- Bogdan, R. J. (1997). *Interpreting minds*. Cambridge, MA: Bradford Books, MIT Press.
- Brafman, R. I., & Tennenholtz, M. (1997). Modeling agents as qualitative decision makers. *Artificial Intelligence*, 217–268.
- Bratman, M. E. (1987). *Intention, plans and practical reason*. Cambridge, MA: Harvard University Press.
- Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4(4), 349–355.
- Bui, H., Venkatesh, S., & West, G. (2002). Policy recognition in the abstract hidden Markov model. *Journal of Artificial Intelligence Research*, 17.
- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, 70(2), 381–395.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3), 861–898.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books, MIT Press.

- Carey, S. (2011). *The origin of concepts*. New York: Oxford University Press.
- Carnap, R. (1942). *Introduction to semantics*. Cambridge, MA: Harvard University Press.
- 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.
- Chajewska, U., & Koller, D. (2000). Utilities as random variables: Density estimation and structure discovery. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI)* (pp. 63–71).
- Chajewska, U., Koller, D., & Ormoneit, D. (2001). Learning an agent's utility function by observing behavior. In *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 35–42).
- Charniak, E., & Goldman, R. (1991). A probabilistic model of plan recognition. In *Proceedings of the tenth national conference on artificial intelligence* (pp. 160–165).
- Charniak, E., & Goldman, R. P. (1993). A Bayesian model of plan recognition. *Artificial Intelligence*, 64, 53–79.
- Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3(2).
- Choi, J., & Kim, K.-E. (2011). Inverse reinforcement learning in partially observable environments. *Journal of Machine Learning Research*, 12, 691–730.
- Churchland, P. M. (1970). The logical character of action-explanations. *Philosophical Review*, 79(2), 214–236.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Cohen, P. R. (1995). *Empirical methods in artificial intelligence*. Cambridge, MA: MIT Press.
- Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, 42, 213–261.
- Cohen, P. R., Morgan, J., & Pollack, M. E. (Eds.). (1990). *Intentions in communication*. Cambridge, MA: MIT Press.
- Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3, 177–212.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*(107), 705–717.
- 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.
- Davis, L. S., & Benedikt, M. L. (1979). Computational models of space: Isovists and isovist fields.

- Computer Graphics and Image Processing*, 11, 49–72.
- 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. (1971). Intentional systems. *Journal of Philosophy*, 68(4), 87–106.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 4, 568–570.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Descartes, R. (1641). *Meditations of the first philosophy*.
- Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B*, 308, 67–78.
- Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13, 227–303.
- Doshi, P., Qu, X., Goodie, A., & Young, D. (2010). Modeling recursive reasoning by humans using empirically informed interactive pomdps. In *Proceedings of the ninth international conference on autonomous agents and multiagent systems (AAMAS)*.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12, 961–974.
- Dvijotham, K., & Todorov, E. (2010). Inverse optimal control with linearly-solvable MDPs. In *Proceedings of the twenty-seventh international conference on machine learning (ICML)* (pp. 335–342).
- Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.
- Feinberg, E. A., & Shwartz, A. (Eds.). (2002). *Handbook of Markov decision processes: Methods and applications*. New York: John Wiley and Sons.
- Feldman, J., & Tremoulet, P. D. (2008). *The attribution of mental architecture from motion: towards a computational theory* (Tech. Rep. No. RuCCS TR-87). Department of Psychology, Rutgers University.
- Filar, J., & Vrieze, K. (1997). *Competitive Markov decision processes*. New York: Springer.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology*, 15(1), 95–120.
- Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, 44(3), 283–296.
- Frank, M. C., Goodman, N. D., Lai, P., & Tenenbaum, J. B. (2009). Informative communication in word production and word learning. In *Proceedings of the thirty-first annual conference of the cognitive science society*.
- Gal, Y., & Pfeffer, A. (2004). Reasoning about rationality and beliefs. In *Proceedings of the third international conference on autonomous agents and multiagent systems (AAMAS)*.
- Gal, Y., & Pfeffer, A. (2008). Networks of influence diagrams: A formalism for representing agents beliefs and decision-making processes. *Journal of Artificial Intelligence Research*, 33, 109–147.
- Geib, C. W., & Goldman, R. P. (2009). A probabilistic plan recognition algorithm based on plan tree grammars. *Artificial Intelligence*, 173(11), 1101–1132.

- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological Review*, 96, 267-314.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd 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., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, 7(7), 287–292.
- Gergely, G., Nádasdy, Z., Csibra, G., & Biró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165–193.
- Gmytrasiewicz, P. J., & Doshi, P. (2005). A framework for sequential planning in multi-agent settings. , 24, 49–79.
- Gmytrasiewicz, P. J., & Durfee, E. H. (1995). A rigorous, operational formalization of recursive modeling. In *Proceedings of the first international conference on multi-agent systems*.
- Gold, E. M. (1967). Language identification in the limit. *Information and Control*, 10, 447–474.
- Goldman, A. I. (2006). *Simulating minds*. New York: Oxford University Press.
- Goldman, R. P., Geib, C. W., & Miller, C. A. (1999). A new model of plan recognition. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI)* (pp. 245–254).
- 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).
- Goodman, N. D., Baker, C. L., & Tenenbaum, J. B. (2009). Cause and intent: Social reasoning in causal learning. In *Proceedings of the thirty-first annual conference of the cognitive science society* (pp. 2759–2764).
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59(1), 26–37.
- Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*(7), 145–171.
- Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116, 661–716.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Com-*

pututional Linguistics, 12, 175–204.

- Grosz, B. J., & Sidner, C. L. (1990). Plans for discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 417–444). 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 Association for Computational Linguistics* (pp. 159–166).
- Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as i do: 7-month-old infants selectively reproduce others' goals. *Developmental Science*, 11(4), 487–494.
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26(1), 30–39.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557–560.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, 13(6), 923–929.
- Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, 4, 576–577.
- Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players: Part I. the basic model. *Management Science*, 14(3), 159–182.
- Harsanyi, J. C. (1968a). Games with incomplete information played by "Bayesian" players: Part II. Bayesian equilibrium points. *Management Science*, 14(5), 320–334.
- Harsanyi, J. C. (1968b). Games with incomplete information played by "Bayesian" players: Part III. the basic probability distribution of the game. *Management Science*, 14(7), 486–502.
- Hauskrecht, M. (2000). Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research*, 13, 33–94.
- 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.
- Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.
- Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, 88, 947–969.
- Hogrefe, G. -Juergen, Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 57(3), 567–582.

- Hume, D. (1739). *A treatise of human nature*.
- James, W. (1890). *The principles of psychology*.
- Jefferys, W., & Berger, J. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, 80, 64–72.
- Jern, A., & Kemp, C. (2011). Capturing mental state reasoning with influence diagrams. In *Proceedings of the thirty-third annual conference of the cognitive science society*.
- Jern, A., Lucas, C. G., & Kemp, C. (2011). Evaluating the inverse decision-making approach to preference learning. In *Advances in neural information processing systems*.
- 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).
- Kelley, H. H. (1972a). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.
- Kelley, H. H. (1972b). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28, 107–128.
- Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, 15, 483–524.
- Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2007). Learning causal schemata. In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.
- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10(3), 307–321.
- Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.
- Körding, K. (1997). Decision theory: What “should” the nervous system do? *Science*, 318(5850), 606–610.
- Krasnow, M. M., Truxaw, D., Gaulin, S. J. C., New, J., Ozono, H., Uono, S., et al. (2011). Cognitive adaptations for gathering-related navigation in humans. *Evolution and Human Behavior*, 32, 1–12.
- Kripke, S. A. (1959). A completeness theorem modal logic. *The Journal of Symbolic Logic*, 24(1), 1–14.
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, 14(5), 402–408.
- Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (p. 121–141). Oxford: Clarendon Press.
- Lewis, C. I., & Langford, C. H. (1932). *Symbolic logic*. Century Company.

- Lewis, D. (1969). Convention: A philosophical study.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.
- 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).
- Litman, D. J., & Allen, J. F. (1987). A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11, 163–200.
- 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).
- Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, 35(4), 549–568.
- Lucas, C., Griffiths, T. L., Xu, F., & Fawcett, C. (2009). A rational model of preference learning and choice prediction by children. In *Advances in neural information processing systems* 21.
- Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, 14(2), 453–460.
- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8), 601–608.
- Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, 105, 489–512.
- Luo, Y., & Baillargeon, R. (2010). Toward a mentalistic account of early psychological reasoning. *Current Directions in Psychological Science*, 19(5), 301–307.
- Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, 12(1), 142–149.
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning and social interaction*. Cambridge, MA: MIT Press.
- 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.
- Milch, B., & Koller, D. (2000). Probabilistic models for agents' beliefs and decisions. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)* (pp. 389–396).
- Mill, J. S. (1874). *Essays on some unsettled questions of political economy* (2nd ed.). London: Longmans, Green, Reader & Dyer.
- Morariu, V. I., Prasad, V. S. N., & Davis, L. S. (2007). Human activity understanding using visibility context. In *Ieee/rsj iros workshop: From sensors to human spatial concepts (fs2hsc)*.
- Neu, G., & Szepesvári, C. (2007). Apprenticeship learning using inverse reinforcement learning and gradient methods. In *Proceedings of the Twenty-third Conference on Uncertainty in Artificial Intelligence (UAI)*.

- Ng, A. Y., & Jordan, M. I. (2000). PEGASUS: A policy search method for large MDPs and POMDPs. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Ng, A. Y., & Russell, S. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 663–670).
- Nichols, S., & Stich, S. (2003). *Mindreading*. Oxford: Clarendon Press.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, 22, 129–151.
- Pareto, V. (1906). *Manual of political economy*.
- 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*. Cambridge, MA: MIT Press.
- Perner, J., Frith, U., Leslie, A. M., & Leekman, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development*, 60(3), 689–700.
- Perner, J., Leekan, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125–137.
- Perner, J., & Wimmer, H. (1985). "john thinks that mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437–471.
- Perrault, C. R., & Allen, J. F. (1980). A plan-based analysis of indirect speech acts. *Computational Linguistics*, 6(3-4), 167–182.
- Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, 98, 137–155.
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant / The origins of intelligence in children*. New York: International University Press, 1952.
- Piaget, J. (1937). *La construction du réel chez l'enfant / The construction of reality in the child*. New York: Basic Books, 1954.
- Piaget, J., & Inhelder, B. (1948). *La représentation de l'espace chez l'enfant / The child's conception of space*. New York: W. W. Norton, 1956.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515–526.
- Puterman, M. L. (2005). *Markov decision processes: Discrete stochastic dynamic programming*. New York: Wiley Interscience.
- Pynadath, D. V., & Marsella, S. C. (2005). Psychsim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the twenty-first international joint conference on artificial intelligence*.
- Pynadath, D. V., & Wellman, M. P. (2000). Probabilistic state-dependent grammars for plan recognition. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Quine, W. V. (1956). Quantifiers and propositional attitudes. *Journal of Philosophy*, 53, 177–187.

- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77.
- Ramachandran, D., & Amir, E. (2007). Bayesian inverse reinforcement learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*.
- Rao, R. P. N., Shon, A. P., & Meltzoff, A. N. (2007). A Bayesian model of imitation in infants and robots. In K. Dautenhahn & C. Nehaniv (Eds.), *Imitation and social learning in robots, humans, and animals*. Cambridge University Press.
- 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.
- Russell, B. (1956). *Logic and knowledge: Essays 1901–1950* (R. C. Marsh, Ed.). London: Unwin Hyman.
- Ryle, G. (1949). *The concept of mind*. New York: Barnes & Noble.
- 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.
- Schmidt, C. F., & Marsella, S. C. (1991). Planning and plan recognition from a computational point of view. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development, and simulation of everyday mindreading* (pp. 109–126). Oxford: Basil Blackwell.
- Schmidt, C. F., Sridharan, N. S., & Goodson, J. L. (1978). The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence*, 11, 45–83.
- 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. *Phil. Trans. R. Soc. Lond. B*, 358(1430), 415–427.
- Schultz, T. R. (1991). From agency to intention: A rule-based, computational approach. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development, and simulation of everyday mindreading* (pp. 79–95). Oxford: Basil Blackwell.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? attributing false beliefs about identity at 18 months. *Child Development*, 80, 1172–1192.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. R. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. E. Pollack

- (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: MIT Press.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4(2), 145–166.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- 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.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*(96), B1–B11.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89–96.
- Stahl, D. O., & Wilson, P. W. (1995). On players models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10, 218–254.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580–586.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.
- Tauber, S., & Steyvers, M. (2011). Using inverse planning and theory of mind for social goal inference. In *Proceedings of the Thirtieth Third Annual Conference of the Cognitive Science Society*.
- 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.
- Todorov, E. (2006). Linearly-solvable Markov decision problems. In *Advances in neural information processing systems 19* (pp. 1369–1376).
- Todorov, E. (Under Review). Efficient algorithms for inverse optimal control. In *International conference on machine learning*.
- 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.
- Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems 22* (p. 1874-1882).
- Vail, D. L., Veloso, M. M., & Lafferty, J. D. (2007). Conditional random fields for activity recognition. In *Proceedings of the sixth international conference on autonomous agents and multiagent systems (AAMAS)*.
- van Benthem, J. (2001). Games in dynamic-epistemic logic. *Bulletin of Economic Research*, 53(4), 219–248.

- van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2006). *Dynamic epistemic logic*. New York: Springer.
- Van Overwalle, F. (2010). Infants' teleological and belief inference: A recurrent connectionist approach to their minimal representational and computational requirements. *Neuroimage*, 52, 1095–1108.
- 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).
- Vygotsky, L. (1934). *Thought and language*. Cambridge, MA: MIT Press, 1986.
- Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy*, 7(1), 73–83.
- Wahl, S., & Spada, H. (2000). Children's reasoning about intentions, beliefs and behaviour. *Cognitive Science Quarterly*, 1, 5–34.
- Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. Unpublished doctoral dissertation, Cambridge University, Cambridge, England.
- Waugh, K., Ziebart, B. D., & Bagnell, J. A. (2011). Computational rationalization: The inverse equilibrium problem. In *Proceedings of the twenty-eighth international conference on machine learning (ICML)*.
- 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.
- Wellman, H. M., Cross, D., & Watson, J. (2001). A meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655–684.
- Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, 43, 337–375.
- Whiten, A. (Ed.). (1991). *Natural theories of mind: Evolution, development, and simulation of everyday mindreading*. Oxford: Basil Blackwell.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Wingate, D., & Seppi, K. D. (2005). Prioritization methods for accelerating MDP solvers. *Journal of Machine Learning Research*, 6, 851–881.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Phil. Trans. R. Soc. Lond. B*, 358, 593–602.
- 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.
- Yoshida, W., Seymour, B., Friston, K. J., & Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *The Journal of Neuroscience*, 30(32), 10744–10751.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28, 979–1008.

- Zettlemoyer, L. S., Milch, B., & Kaelbling, L. P. (2009). Multi-agent filtering with infinitely nested beliefs. In *Advances in neural information processing systems 21*.
- Ziebart, B. D., Bagnell, J. A., & Dey, A. K. (2010). Modeling interaction via the principle of maximum causal entropy. In *Proceedings of the twenty-seventh international conference on machine learning (ICML)*.