Cognitive Psychology 123 (2020) 101334

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

The Naïve Utility Calculus as a unified, quantitative framework for action understanding

Julian Jara-Ettinger[a,b,*], Laura E. Schulz[c,d], Joshua B. Tenenbaum[c,d]

[a] Department of Psychology, Yale University, United States
[b] Department of Computer Science, Yale University, United States
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States
[d] Center for Brains, Minds and Machines, Massachusetts Institute of Technology, United States

ARTICLE INFO

Keywords:
Action understanding
Bayesian models of cognition
Social cognition
Theory of mind

ABSTRACT

The human ability to reason about the causes behind other people' behavior is critical for navigating the social world. Recent empirical research with both children and adults suggests that this ability is structured around an assumption that other agents act to maximize some notion of subjective utility. In this paper, we present a formal theory of this Naïve Utility Calculus as a probabilistic generative model, which highlights the role of cost and reward tradeoffs in a Bayesian framework for action-understanding. Our model predicts with quantitative accuracy how people infer agents' subjective costs and rewards based on their observable actions. By distinguishing between desires, goals, and intentions, the model extends to complex action scenarios unfolding over space and time in scenes with multiple objects and multiple action episodes. We contrast our account with simpler model variants and a set of special-case heuristics across a wide range of action-understanding tasks: inferring costs and rewards, making confidence judgments about relative costs and rewards, combining inferences from multiple events, predicting future behavior, inferring knowledge or ignorance, and reasoning about social goals. Our work sheds light on the basic representations and computations that structure our everyday ability to make sense of and navigate the social world.

1. Introduction

> "… a man, being just as hungry as thirsty, and placed in between food and drink, must necessarily remain where he is and starve to death."
> — Aristotle, On the Heavens 295b, c. 350 BCE

People naturally interpret each other's behavior by attributing mental states such as beliefs, desires, and intentions. If, for instance, someone picks up their mug and immediately puts it back down, we can infer that they *wanted* to drink what they *thought* was in the mug, and that they *realized* that the mug was empty when they picked it up. These mental-state inferences help us explain why people act the way they do, predict what they'll do next, and they guide us in how to react: If we think our friend wanted coffee, we might expect her to get up and walk to the kitchen, mug in hand. If we just took the last cup, we might try to be helpful and point out we're out of coffee as soon we see her get up, even without direct evidence that she was, in fact, planning to get some more.

Although the ability to attribute beliefs and desires develops throughout childhood and into adolescence (Wellman, Cross, &

* Corresponding author.
E-mail address: julian.jara-ettinger@yale.edu (J. Jara-Ettinger).

https://doi.org/10.1016/j.cogpsych.2020.101334
Received 6 July 2019; Received in revised form 22 May 2020; Accepted 17 July 2020

Watson, 2001; Richardson, Lisandrelli, Riobueno-Naylor, & Saxe, 2018), its building blocks are at work from early in infancy. Before their first birthday, infants already interpret other people's actions as meant to complete goals (Woodward, 1998; Woodward, Sommerville, & Guajardo, 2001; Skerry, Carey, & Spelke, 2013), and they infer these goals by assuming that agents move efficiently in space (Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra, & Bíró, 1995;, Csibra, Gergely, Bíró, Koós, & Brochkbank, 1999; Csibra, Bíró, Koós, & Gergely, 2003; Scott & Baillargeon, 2013). Yet, adult commonsense psychology goes far beyond goal-attribution.

To illustrate this, consider a simple example from a moment in childhood many of us are all too familiar with. Suppose that you are in preschool, and you and your classmate each ask the teacher for help at the same time. The teacher looks at each of you briefly and then walks towards your classmate. Your teacher's goal is clearly to help your classmate. But there are many ways to explain the causes behind this goal. Perhaps the teacher likes your classmate better. More likely, your classmate just happened to be closer, or louder. The teacher might know that what you need can wait. Or they may be confident that you don't need help, even if you think you do. Each of these explanations boils down to the same goal—to help your classmate—but each explanation licenses different expectations about what the teacher will do next, and it affects how we evaluate their actions. A teacher who decides not to help you based on their personal preferences, for instance, has a different moral standing than a teacher who decides not to help you because they want you to challenge yourself.

How do we represent and infer the causes behind other people's goals? Research suggests that inferences beyond goal attribution are supported by an expectation that agents make choices by quantifying, comparing, and maximizing subjective utilities—the difference between the costs they incur and the rewards they obtain. This *Naïve Utility Calculus* allows us to infer the knowledge, preferences, and moral values that explain other people's goals (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Jern, Lucas, & Kemp, 2017; Kleiman-Weiner, Saxe, & Tenenbaum, 2017; Lucas et al., 2014), and empirical work suggests that even young children share these expectations (Jara-Ettinger, Tenenbaum, & Schulz, 2015; Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017; Pesowski, Denison, & Friedman, 2016; Lucas et al., 2014), with some basic form of the Naïve Utility Calculus in place in infancy (Liu, Ullman, Tenenbaum, & Spelke, 2017).

Despite robust empirical support for the Naïve Utility Calculus, a number of critical questions remain open. All of these questions reflect aspects of a single overarching concern: To what extent is this "naïve utility calculus" really best thought of as a "calculus"—a coherent, unified, quantitative and rational inferential framework? We focus on three specific aspects here. First, all studies to date reveal either the agent's costs and ask people to infer the rewards, or vice-versa. In realistic scenarios, we often know neither, so we must jointly infer the costs and rewards from how an agent acts on one or more occasions. Does the Naïve Utility Calculus represent cost-reward tradeoffs with a coherent generative model of action that can support these more complex joint inferences? Second, does the Naïve Utility Calculus operate only on coarse representations of costs and rewards, supporting only qualitative inferences, or can it also drive fine-grained quantitative inferences by tracking exact tradeoffs between costs and rewards? Finally, and most generally, is the Naïve Utility Calculus best thought of as a unified generative model of how agents act given different costs and rewards, supporting many different judgements via mechanisms of approximate probabilistic inference, or instead as a more piecemeal collection of simple, cheap heuristics that only approximate rational inferences in special cases?

We answer these questions by formalizing the Naïve Utility Calculus in a computational model that performs approximate Bayesian inferences of costs and rewards over extended sequences of actions that unfold over time and space. Our model builds on but extends substantially beyond previous qualitative formulations (Jara-Ettinger et al., 2016; Jara-Ettinger, Floyd, Huey, Tenenbaum, & Schulz, 2019; Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017), as well as simpler quantitative formulations of utility-based action understanding (e.g., Baker et al., 2017; Lucas et al., 2014; Jern, Lukas, & Kemp, 2017) that do not attempt to account for inferences about multiple dimensions of cost and reward, or complex actions operating over multiple spatial and temporal scales. We then present a set of quantitative experiments that test if (1) the Naïve Utility Calculus supports joint inferences of costs and reward from observable actions; if (2) these inferences can be captured with quantitative precision; and (3) if these judgments are best explained by a unified theory structured around the single assumption that agents approximately maximize utilities. Throughout these studies, we compare our full Naïve Utility Calculus model with a number of variants ablating different aspects of the model, as well as simpler accounts that make similar qualitative but different quantitative predictions. These findings provide more direct evidence for each of the model's main components, and suggest that action-understanding in people's intuitive psychology is structured as a coherent, causal generative model of agents' actions and choices, rather than as collection of special-purpose inference heuristics.

1.1. The Naïve Utility Calculus: An informal overview

We begin by defining the notion of utility we will work with. When people observe intentional behavior, we assume that they attempt to understand it as implementing a plan intended to achieve some outcome, and that plan is chosen according to a subjective utility function

$$U(p, o) = \mathrm{R}(o) - \mathrm{C}(p). \tag{1}$$

Here $U(p, o)$ represents the utility expected from acting according to plan p expecting to successfully reach outcome o, which in its most basic form can be expressed as the difference between $\mathrm{R}(o)$, the subjective reward the agent expects to receive from that outcome, and $\mathrm{C}(p)$, the subjective cost that the agent expects to incur in executing the plan.

At the heart of the Naïve Utility Calculus is the assumption that agents act to maximize this utility function. That is, agents decide how to act by effectively estimating the utility associated with different action plans and pursuing the one that yields the highest utility. We assume, crucially, that agents stochastically estimate their subjective utilities rather than knowing the precise values. Our model is consequently approximate and probabilistic: Agents select the action plan with the highest estimated subjective utility, but

this does not always correspond to the action plan with highest true subjective utility. Through this assumption, we can model how people infer the cost and rewards behind other people's actions as Bayesian inference, positing the configuration of costs and rewards that explains the observed actions.

Here we focus on the Naïve Utility Calculus in the context of agents moving in space (such as those shown in in Fig. 3) and we use concrete notions of costs and rewards that even young children can grasp (costs associated with physical actions, and rewards associated with reaching for objects or helping agents; Woodward, 1998; Liu et al., 2017; Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017). This allows us to perform quantitative tests of our account in basic settings without relying on linguistic information about agents' behaviors, but the inferences that we formalize can generalize to more abstract notions of cost and reward. We return to this point in the discussion.

This Naïve Utility Calculus is an account for how people intuitively make sense of other people's behavior, and is not meant to imply any assumption that people actually compute (let alone maximize) utilities when they act. Violations of classical utility theory in human behavior are well known: People are averse to risky choices, even when they have higher expected utilities (Kahneman & Tversky, 1979); they do not update their utility estimates appropriately (Vos Savant, 1990); and some patterns of choices cannot be explained through utility functions (Allais, 1953). These and other violations of maximal expected utility decision making are important but do not in any way compete with our claims here, which are about the *intuitive theory* of decision making. Indeed, these failure cases in a way support our theory. The fact that nonexperts are often surprised when confronted with phenomena showing utility theory failing as a descriptive account of decision making is exactly what we would expect if our intuitive theory of others assumes that agents maximize utilities.

Having said this, we do expect that at least in the most basic cases of human action where our intuitive theories are well developed and applied on a routine basis—for instance, making sense of people reaching for objects around them, or navigating through space in their immediate environment to reach goals—people's choices should be to some degree approximately utility maximizing. If utility theory were entirely inapplicable in these cases, it would not have any explanatory or predictive power as an intuitive theory. Utility theory enjoyed enormous success in classical economics precisely because it appears to explain human behavior approximately and intuitively in many basic, everyday situations (Von Neumann & Morgenstern, 1944; Brown, 1986). Our goal here is to assess formally how and whether a simple version of utility theory, embedded in a Bayesian framework, can provide a strong quantitative account of how naïve human observers infer other agents' costs and rewards from their actions.

Before presenting the Naïve Utility Calculus model formally, we walk through a set of basic predictions and assumptions underlying the account, motivated by qualitative phenomena of action understanding that can be observed in both adults and young children. We then present the model and a number of alternatives in quantitative terms, followed by a series of behavioral experiments rigorously testing these accounts against each other.

Fig. 1 shows the basic workings of the Naïve Utility Calculus in simple situations that even infants understand (Liu & Spelke, 2017; Gergely & Csibra, 2003). If we assume that agents maximize utilities, then we must expect agents to only act when the rewards outweigh the costs (otherwise, not acting at all yields a higher utility; Fig. 1a). When agents do act, the expectation that agents maximize utilities implies that they will fulfill their goals as efficiently as possible (because lower costs yield higher utilities; Fig. 1b). Fig. 1c illustrates how the Naïve Utility Calculus supports inferences about the causes behind other people's goals. In Figs. 1c-1 and 1c-2, both agents clearly chose the purple star over the green star. But, intuitively, the agent in Fig. 1c-1 revealed their preference more clearly relative to the agent in Fig. 1c-2. This is predicted by the Naïve Utility Calculus. The agent in Fig. 1c-1 incurred a higher cost to obtain the purple star, which can only be explained by positing a higher reward. By contrast, the agent in Fig. 1c-2 incurred a low cost, which is consistent with the agent having a weak or a strong preference for the purple star. Figs. 1c-2 and 1c-3 show the opposite contrast: Although both agents chose the purple star over the green star, the agent in Fig. 1c-2 reveals a preference, while the agent in 1c-3 does not. This again is predicted by the Naïve Utility Calculus: The agent in Fig. 1c-3 may have preferred the green star, but not enough to be willing to jump over the obstacle to get it. Although the qualitative formulation of the Naïve Utility Calculus is relatively simple, a wide range of intuitions in action-understanding can be explained by it (see Jara-Ettinger et al., 2016 for a review of qualitative implications of the Naïve Utility Calculus and their relation to developmental research; see also Lucas et al., 2014; Jern & Kemp, 2011, 2014; Jern, Lucas, & Kemp, 2011, 2017).

Because different people have different preferences and abilities, the Naïve Utility Calculus can only yield reasonable inferences if costs and rewards are treated as agent-dependent. This is illustrated in Fig. 1d-e. In Fig. 1d, although one agent takes a physically longer path, we do not conclude that she failed to maximize utilities. Instead, we infer that one agent finds jumping easier than walking around the wall, whereas the other agent does not. Similarly, the event in Fig. 1e cannot be explained if we assumed that both agents have identical subjective rewards. Instead, by assuming that agents maximize utilities, we infer that the blue agent prefers the orange star whereas the purple agent may have chosen the green story only because it was closer (and even toddlers recognize that different agents can have different rewards; Repacholi & Gopnik, 1997; Doan, Denison, Lucas, & Gopnik, 2015; Ma & Xu, 2011).

As noted above, these intuitions trace back to early childhood. By age five, children's reasoning about the costs and rewards behind other people's goals suggest they assume agents maximize utilities (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). Moreover, when an agent fails to maximize their true utilities, children infer that the agent must have been ignorant about her own costs or rewards, suggesting that by age four we already understand that agents maximize the utilities they expect to obtain, and not the true utilities they obtain (Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017). At an even earlier age, inferences of this kind already play a role in social evaluations. Two-year-old toddlers judge a competent agent who refuses to help more harshly relative to a less competent agent who also refuses to help (Jara-Ettinger et al., 2015).

The developmental studies are useful in establishing the degree to which reasoning about agents' costs and rewards is

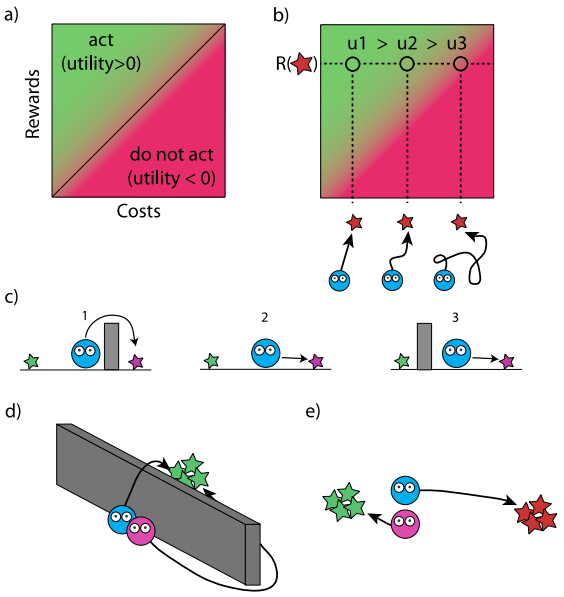

Fig. 1. (a) relation between a goal's costs and rewards. The green region represents the space of positive utilities (where the rewards outweigh the costs) and the red region represents the space of negative utilities (where the costs outweigh the rewards). Assuming that agents maximize utilities implies that only positive utilities can motivate agents to act. This simple schematic reveals some direct implications of using a Naïve Utility Calculus to reason about others. When the cost is low (left side of the figure), a wide range of rewards produce positive utilities. Thus, when we watch someone pursue a low-cost goal, we cannot infer her reward (both high and low rewards produce positive utilities). By contrast, when the cost is high (right side of the figure), only high rewards can produce positive utilities. Thus, when agents pursue a high-cost goals, we can infer that the reward was even higher. The structure of these inferences flips in the case of inaction. If an agent foregoes a low-cost goal, we can be sure that the reward was even lower. However, if an agent foregoes a high-cost goal, we may be unsure about whether the reward was also low, or if it was high, but not high enough. (b) More efficient paths are, by definition, less costly, and therefore produce higher utilities. Thus, the expectation that agents maximize utilities implies an expectation that agents move efficiently in space. (c) Examples of graded preference inferences. In all cases the agent's goal is the same, but the inferred preference changes depending on the relative costs. (d) Costs vary across agents. When two agents pursue the same goal, but take different actions, we do not infer that one of them failed to maximize utilities. Instead, we infer that the two agents have different subjective costs. In this case, one agent finds it easier to jump over the wall than the other. (e) Rewards vary across agents. When two agents pursue different goals, we infer that the two agents have different subjective rewards. In this case, the agent who traveled to the farther rewards reveals a preference while the other agent does not. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

foundational and emerges early in social cognition (providing evidence for the "naïve" in Naïve Utility Calculus). Our principal goal in this work is to develop a more quantitative computational model of the Naïve Utility Calculus and test it rigorously—thus providing evidence that the "Naïve utility calculus" really is best thought of as a "calculus." We work within the framework of probabilistic generative models (Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Goodman, Tenenbaum, & Gerstenberg, 2014), applied to how agents make decisions and plan actions. We formalize a generative model of planning that takes as its input an agent's subjective costs and rewards, together with situational information (e.g., the location of different objects and terrains) and, through the assumption of utility maximization, determines the agent's goals and actions (see Fig. 2). Given this generative model, an observer can apply Bayesian inference to approximately invert the planning process and infer the costs and rewards most likely underlying an agent's behavior.

Our work builds on a number of important earlier computational proposals. A first family of models, known as *inverse decision-making*, has focused on formalizing how we infer preferences behind people's choices (Lucas et al., 2014; Jern et al., 2017; Jern & Kemp, 2014). These models also work through an assumption that agents maximize utilities. However, inverse decision-making models have only been used to test preference inferences from isolated, discrete choices (e.g., choosing between eating fish or turkey; Jern et al., 2017) rather than events with complex spatiotemporal structure like the ones we study here. In addition, these models have only been tested in situations where costs are not involved and only rewards need to be inferred. By contrast, our model uses utility-based reasoning about agents with variable costs and rewards, taking extended sequences of actions with multiple goals unfolding over time and space. This allows us to test the key hypotheses about the Naïve Utility Calculus that previous models could not answer: Can people perform joint inferences over costs and rewards? Are these inferences quantitative and fine-grained, in ways that respond to the precise structure of the spatial environment and the temporal sequence of actions? And are the rich patterns of inference that can be studied in these complex action settings all coherently explained by a single unified probabilistic generative model?

Our model is more closely related to a second family of models, known under the umbrella term *inverse planning*. These models formalize action understanding through Markov Decision Processes (Baker, Saxe, & Tenenbaum, 2009; Baker, Jara-Ettinger, Saxe, &

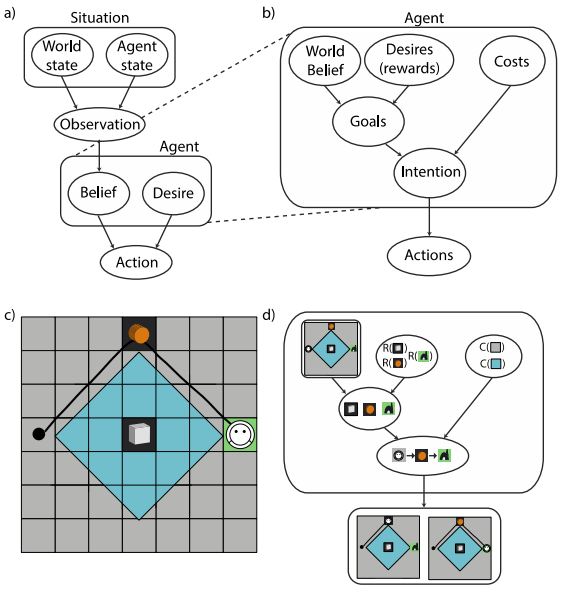

Fig. 2. Sketch of the generative model. (a) Previous model of goal-directed action understanding (Baker et al., 2017). Under this generative model, the world and the agent's location determine what the agent can see; these observations change the agent's beliefs rationally; and the beliefs, combined with the desires, determine the agent's actions. (b) Generative model of the Naïve Utility Calculus as a refinement of the agent's internal structure. Our model distinguishes hierarchically between goals (states of the world that the agent finds rewarding), intentions (ordered sequences of goals represented as action plans with estimated costs and rewards), and actions (a compilation of an intention into an action policy). The expectation that agents maximize utilities determines which intention is selected, and how it is completed. (c) Schematic of stimulus from our paradigm. An astronaut in an alien planet travels from the middle left spot of the map to the space station on the middle right side of the map. The astronaut walks around the blue terrain and stays on the gray one. She also collects an orange (cylindrical) care package but foregoes the white (cubic) one. (d) A concrete application of the model schematic from (b). Given a set of costs and rewards, the agent forms an intention—an ordered sequence of goals—that maximizes their utilities. In this case, the intention is to obtain the orange cylinder, and to then go to the space station. Each of these goals is then completed by executing the policy from a goal-specific MDP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Tenenbaum, 2017; Jara-Ettinger, Sun, Schulz, & Tenenbaum, 2018; Ullman et al., 2009), which does allow them to address actions with spatiotemporal structure. While our model also embodies a form of inverse planning, it differs from past models in two significant ways. First, previous models do not emphasize or attempt to explain the multiple causes behind other people's goals. Our model in contrast integrates both expectations about how agents navigate towards their goals, and expectations about how agents choose which goals to pursue in the first place. We achieve this by distinguishing between agents' desires, goals, intentions, and actions, providing a more expressive picture of how people represent about other people's minds. A second difference between our model and past inverse planning models is that our model focuses on the problem of jointly inferring agent-specific costs and rewards from observable actions. By contrast, past work treated costs as constant, observable, and uniform across agents, making them unsuitable for testing the key hypotheses our work aims to test. Our formulation allows us to explain how people jointly infer the costs and rewards that interact to determine an actor's goals; how people predict the ways agents will act in new situations when environmental affordances (and corresponding costs and rewards) vary; how people reason about agents who are still learning about their environments, and learning their own costs and rewards over time; and how people make social evaluations by appealing to the costs agents choose or refuse to incur when deciding whether to help another—all of which previous models are unable to handle.

Our experiments were designed to test the model's predictions about these cases, specifically in response to our three critical questions introduced earlier. To answer our first two questions—Does the Naïve Utility Calculus support joint cost and reward inferences from observable actions? If so, are these fine-grained quantitative inferences tracking exact tradeoffs between costs and rewards? —we presented participants with an agent acting in a novel world and we asked them to jointly infer the agent's cost and the reward functions. We then compared the functions that participants provided with the ones that our model inferred (Experiment 1). We next tested if our model continues to quantitatively predict participant judgments in situations where behavior from multiple events needs to be combined to draw the appropriate cost and reward inferences (Experiment 2). Finally, we tested if people's confidence in their inferences matches the confidence in our model (Experiment 3). Combined, these three experiments provide support that our model predicts participant judgments with fine-grained accuracy.

To answer our third question—is the Naïve Utility Calculus instantiated as a probabilistic generative model of how agents act given different costs and rewards? —we presented participants with a variety of action-understanding tasks that can be solved with the same generative model. We then compared participant judgments against the unified predictions of the Naïve Utility Calculus, and against a collection of simple special-case heuristics. Specifically, we tested how participants predict what an agent will do next (Experiment 4), how participants infer whether an agent is knowledgeable or ignorant (Experiment 5), and how participants make

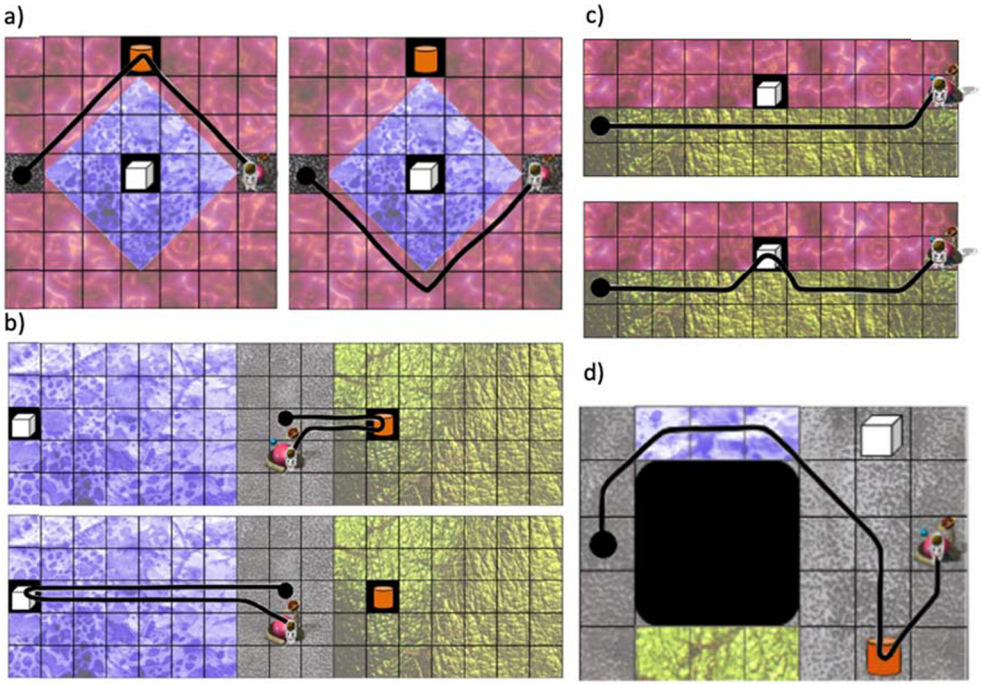

Fig. 3. Examples of experimental stimuli. (a) Example trials used in Experiment 1a where people were asked to infer the astronaut's costs and rewards. (b) Example trials used in Experiment 1b. (c) Example trials used in Experiments 1c. d) Example trial used in Experiment 5. For the experiments, we generated up to twelve different version of each stimulus (by varying the colors of the terrain and colors of the care packages; see Appendix for different versions of each stimulus) and participants were shown a random one for each of the sixteen trials.

simple social evaluations (Experiment 6). We find that the Naïve Utility Calculus model outperforms simple heuristics by predicting fine-grained structure in participant judgments that the heuristics predict should be noise. All experiments were run with naïve participants who had not participated in any of our previous tasks.

2. Computational framework

Our computational model (code available at http://www.github.com/julianje/bishop) consists of two components. The first is a generative model that, given a cost and a reward function, probabilistically produces utility-maximizing behaviors. The second is a mechanism that uses the generative model to recover the costs and rewards underlying an agent's observable actions via Bayesian inference (or more precisely, a Monte Carlo sampling-based approximation to Bayesian inference known as likelihood weighting; see Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Shi, Griffiths, Feldman, & Sanborn, 2010 for an overview of this approach).

Implementing the Naïve Utility Calculus requires formalizing the ingredients of intentional action and choice that make up the utility computations in Eq. (1): notions of plans and outcomes, and their associated costs and rewards. Here we adopt the standard engineering formalism of probabilistic state-space planning, where a plan comprises a sequence of actions (or a policy for generating those actions) that is expected to take the agent through a sequence of states of the world, and an intended outcome is the target sequence of states and actions that the agent is planning to achieve. In many simple settings for planning, costs are associated with each action taken and rewards are associated with goal states: the final state in the outcome intended from a given plan. But in general, both costs and rewards may depend on both states and actions: A plan may be costly because it is expected to pass through states the agent finds aversive or painful, or because the actions needed to traverse states are intrinsically difficult or otherwise costly; end goals may be the most typical source of rewards but certain actions or intermediate states as subgoals may themselves also be intrinsically rewarding. In the remainder of the paper we thus adopt a formalism where both cost and reward functions are defined over combinations these lower-level action and state primitives, and we define specific cost and reward functions on state-action pairs that capture the intuitive notions of varying costs of plans and rewards for intended outcomes appropriate in different behavior contexts.

Our model builds on other utility-based models of action understanding (e.g. Baker et al., 2017; Lucas et al., 2014) by fleshing out the representation of how people intuitively understand agents as making choices and planning actions. We do this through a hierarchical representation spanning four levels (Fig. 2): desires (formalized as a reward function), goals (formalized as states of the world that the agent believes yield rewards; e.g. believed locations of objects), intentions (formalized as ordered sequences of goals intended to maximize utilities), and actions (mappings from states to actions which, when executed, sequentially fulfill each goal in the intention). Only the final lowest level of actions is directly observable to others. Under this framework, action understanding consists of observing other people's actions, and using the assumption of utility maximization to infer the three higher levels of this

hierarchy (desires, goals, and intentions), as well as the underlying costs and rewards. The hierarchical representation supports the understanding that agents can have many desires which the agent may forego because the costs are too high, because the rewards are not high enough, or because of a combination of both. Furthermore, it enables observers to distinguish between a number of different abstract states of mind, such as wanting an object (assigning it a high utility in the moment) and liking an object (assigning it a high reward in general).

2.1. Generative model

Our formal model builds on a general architecture for goal inference based on a principle of efficiency shared with recent proposals such as Baker et al. (2017) and Lucas et al. (2014), and inspired by the framework theories of mind that developmental psychologists have posited to explain children's action explanations (Wellman 1990; 2014). Fig. 2a illustrates this architecture, and Fig. 2b shows how our present model extends this approach to reason about the specific roles that costs and rewards play in the hierarchical computations leading ultimately to agents' actions: rewards combined with beliefs determine candidate goals, and these goals integrated with action costs give rise to specific intentions to act. Fig. 2c-d illustrate this model applied to one of the scenarios that we test in our experiments. We adopted scenarios designed to minimize the role of any specific prior knowledge participants might bring to the task, and instead to draw only on the general schemas of rational action planning formalized in the Naïve Utility Calculus. In this paradigm, an astronaut (shown from a bird's eye view) navigates an alien planet that has different types of terrains (Fig. 3a) that may be more or less exhausting to travel through. The astronaut ultimately wants to reach her space station, but there also one or more care packages that the astronaut can pick up if she wishes. Given the astronaut's actions, the task is to infer how exhausting the astronaut finds it to cross different terrains (the costs), and how much she likes each care package (the rewards).

In explaining our generative model, we find it most natural to begin in the bottom of the planning hierarchy, by describing how planning turns intentions into action policies. We then explain how the model builds intentions by computing and maximizing utilities.

2.1.1. From intentions to action policies

We define intentions as an ordered sequence of goals the agent plans to pursue, where each goal corresponds to reaching a target position in the map. In Fig. 1c-d, for instance, the agent's intention is to collect the orange care package and to go to the space station, in that order. An intention's utility is given the sum of rewards the agent obtains in completing each goal, minus the total cost the agent incurs to complete these goals (Eq. (1)).

We determine how agents complete goals (and, consequently, the cost associated with completing each goal) through Markov Decision Processes (MDPs), a standard mathematical framework in AI applications (Sutton & Barto, 1998) for describing how agents can act to maximize expected rewards in simple worlds. While past work has used MDPs to model an agent's full planning process (Baker et al., 2009), here we use them in a more restricted way. In our framework, MDPs serve to determine how to reach specific goals. Consequently, each potential goal is associated with its individual special-purpose MDP that uses a single reward associated with the target goal rather than having access to all of the agent's rewards. Thus, MDPs are used to plan movement in space, but do not play a role in deciding which goals to pursue.

In MDPs, the world is represented as a set of possible states, with each state capturing the physical environment and the agent's location in space S. In a 7x7 world (like the one shown in Fig. 2c), for example, the state space consists of 49 states, each capturing the agent's position in the map. At each discrete time step, the agent can take one of eight actions: move in one of the four cardinal directions, or move in one of the four diagonal directions. Although in MDPs actions can change the world probabilistically, here we assume that actions deterministically move agents in their intended direction (except on map borders, where actions that would take the agent off the map do not change the state). Whenever the agent takes an action, they incur a cost and they may obtain a reward.

We define cost functions, C: $A \times S \rightarrow \Re$, as mappings from combinations of actions and states onto real values. All actions incur a cost determined by the type of terrain the agent is in (e.g., in Fig. 2c, the astronaut finds all gray terrain to be equally costly to navigate through). Formally, $C(a, s_a) = C(a, s_b)$ for all $a \in A$, whenever s_a and $s_b$represent states where the agent is in the same type of terrain. Because the four diagonal actions produce longer movements, $C(a_1, s_a) = \sqrt{2}\,C(a_2, s_a)$, when a_1 is a diagonal action and $a_2$is not (where $\sqrt{2}$ is derived from the Pythagorean theorem).

We define reward functions, R: $A \times S \rightarrow \Re$, as mappings from combinations of actions and states onto real values. Only the MDPs goal state has a reward different from 0. For instance, in the map in Fig. 1c, the MDP responsible for determining how to get to the space station, has a cost associated with each terrain, and a single reward in the space station is (on the middle right end of the map). Similarly, in the same map, the MDP responsible for determining how to get to the white care package has the same costs as the previous MDP, but now with a single reward where the white care package is located.

Using this formulation, we can compute a function that maps states to actions, π: $S \rightarrow A$, called a policy, which, when pursued, reaches the goal state as efficiently as possible. In the MDP framework, this policy can be computed through several different algorithms, any of which converges to the same answer. The Naïve Utility Calculus requires observers to be able to reason about efficient goal-directed actions, but it does not place any constraints on the specific algorithm that must be used to compute efficiency. Here we use value iteration, which constructs the policy by solving the recursive equation

$$V^*(s) = max_a \gamma \sum_{s'} P_{s,a}(s')V^*(s') + R(a, s) - C(a, s) \tag{2}$$

where $V^*(s)$ is a state's optimal value, $P_{s,a}(s')$ is the probability that the agent will be in state s' when she takes action a in state s, and

$\gamma \in (0, 1)$ is a discount parameter that guarantees that the equation converges (see Sutton & Barto, 1998 for a detailed tutorial on MDPs and value iteration). Intuitively, the value function (Eq. (2)) assigns each state a value corresponding to the goal-specific utility (the difference between the rewards and costs) that agents obtains, plus the expected future value, provided that the agent will take the best possible action. Having computed each state's value, the optimal policy is given by

$$\pi^*(s) = argmax_a \left\{ \sum_{s'} P_{s,a}(s') V^*(s') \right\} \tag{3}$$

That is, the optimal policy maps states to actions by maximizing expected value. Because people do not necessarily expect other agents to always act optimally, we relax the assumption that agents act optimally (Eq. (3)) to an expectation that agents are more likely to choose good, over bad, actions. Formally:

$$p(a|s) \propto \exp\left(\sum_{s'} P_{s,a}(s') V^*(s') / \kappa_a \right) \tag{4}$$

This decision rule is known as softmaxing or a Boltzmann policy. $\kappa_a \in (0, \infty)$ is a parameter that determines the agent's "rationality." When κ_a is close to 0, the policy selection becomes optimal and the agent always takes the best action (i.e. Eq. (4) converges to Eq. (3)). As κ_a increases, the policy selection becomes noisier (i.e. Eq. (4) converges to a uniform distribution over actions). Related work has attempted to characterize how the parameter κ_a relates to people's psychological intuitions about rationality (Kryven, Ullman, Cowan, & Tenenbaum, 2015). Because this is not the focus of our model, we did not fit this parameter to participant judgments and we set it to 0.1. For a value this low, κ_a has minimal influence on our model results, but, as we describe in the next section, it enables our model to infer the underlying costs and rewards with fewer samples than would otherwise be required.

Finally, having a set of MDPs, each associated with a goal in the agent's intention, fulfilling the intention reduces to executing the goal-specific MDPs for each goal until all of them are complete.

2.1.2. From desires to intentions and goals

Having defined how agents transform intentions into actions, we next describe how agents form their intentions given their desires.

At the highest level, we formalize desires as rewards associated with having different objects and/or helping different agents. These rewards, combined with the location of the objects or agents who need help determines the space of possible goals. The event in Fig. 1c, for instance, has three goals: collect the top care package, collect the middle care package, and go to the space station. As described above, each of these goals is associated with its own individual MDP-based planner that transforms the goal into actions (see section above): one for reaching the top object, one for reaching the middle object, and one for reaching the space station.

The space of goals next determines the agent's space of intentions, consisting of ordered sequences of goals. Rather than considering every possible intention, our model only considers those that satisfy certain context-specific constraints (formalized as simple mappings from the space of intentions onto True/False values). For instance, in the scenario in Fig. 2b, requiring that the astronaut end in the space station can be implemented as a constraint that only considers intentions where this space station is the final goal. In the event in Fig. 1c, the application of this constraints results in five possible intentions: "go to space station", "get top care package and go to space station", "get middle care package and go to space station", "get top care package, get middle care package, and go to space station", and "get middle care package, get top care package, and go to space station."

Each intention's utility is given by the sum of rewards associated with each goal, minus the sum of the costs that the agent needs to incur to complete the goals in the specified order (Eq. (1)). We assume that agents select an intention through probabilistic utility maximization, such that the probability of selecting intention I is given by

$$p(I) \propto \exp\left(\frac{U(I)}{\kappa_c} \right). \tag{5}$$

Thus, our model applies a graded expectation for utility maximization at an action level (softmaxed action selection in Eq. (4)) and at a choice level (softmaxed intention selection in Eq. (5)). This probabilistic utility maximization at the choice level gives the model a graded expectation over relative utilities.[1] As with the parameter κ_a (Eq. (4)), the parameter κ_c (Eq. (5)) was also not fit to participant judgments and it was set to 0.01. Our model architecture and publicly available code allow easy exploration of how expectations for rational action and rational choice influence mental state inferences in the Naïve Utility Calculus.

After the intention is selected through Eq. (5), actions are produced by completing each goal sequentially, with each goal executed under its autonomous controller specific to that goal solved through an MDP (see from intentions to action policies section).

2.1.3. Summary of generative model

The complete planning process, from desires to actions, is illustrated in Fig. 2c-d. In Fig. 2c, the astronaut initially lands on the

[1] For instance, if two intentions have utilities 10 and 9, respectively, the model will expect the agent to be slightly more likely to pursue the first intention over the second one. By contrast, if the plans have intentions 10 and 1, the model will expect the agent to pursue the first intention over the second one. A model that deterministically selects the highest utility would always select the first intention and would be insensitive to the relative utilities of competing plans.

middle left side of the map and she has to make her way to the space station at the middle right side of the map. There is a white care package in the middle of the map, and an orange care package in the top middle area of the map. Here, the astronaut navigated around the blue terrain and she collected the contents from the orange care package. Fig. 2d shows how this behavior is explained by the generative model. The agent has perfect knowledge about the location of the care packages, the space station, and the terrains, so her world belief corresponds to the true state of the world. The agent obtains rewards by collecting the care packages, or reaching the space station. Thus, the space of goals corresponds to the states of the world where the agent can collect the care packages or where the space station is. These goals, combined with the costs, determine the astronaut's intention. Once the intention has been selected, the action policy associated with each goal is executed sequentially.

To summarize, any state in the world where the agent may obtain a reward is associated with a potential goal. This space of goals, combined with context-specific constraints (such as being able to pursue one goal, or having to always complete a specific goal at the end), determines the space of intentions. Each intention is an ordered sequence of goals, and its utility is given by the sum of the goal's rewards minus the costs that the agent would incur to complete these goals in the specified order (Eq. (1)). The costs are computed through a set of goal-specific MDPs that determine what actions the agent should take to achieve an individual goal (Eqs. (2)–(4)). An intention is then selected through probabilistic utility maximization (Eq. (5)), and this final intention is transformed into actions by executing the action policy (Eq. (4)) associated with each goal (obtained through each goal's MDP).

2.2. Inference over the generative model

Our generative model transforms cost and reward functions into actions. Normally only actions are observable, and costs and rewards must be inferred. We posit that people can perform (approximate) Bayesian inference over this model to infer the unobservable costs and rewards given a sequence of observable actions. This approach has proved successful in past models of social reasoning (e.g. Baker et al., 2009; 2017; Lucas et al., 2014; Verma & Rao, 2005; Hawthorne-Madell & Goodman, 2015). Given a set of observed actions, the posterior probability of the cost and reward functions is given by

$$p(C, R|A) \propto l(A|C, R)p(C, R) \tag{6}$$

where $p(C, R)$ is the prior probability over agents' cost and reward functions and $l(A|C, R)$ is the likelihood that an agent with given costs and rewards would take the observed actions. In our design, we use novel terrains and objects, allowing us to use uniform priors. The likelihood term is given by the probability that the generative model would generate the observed actions, given the costs and rewards. Formally:

$$l(A|C, R) = \sum_{I \in Intentions} p(A|I)p(I|C, R) \tag{7}$$

where $p(I|C, R)$ is the probability that the agent has unobservable intention I given the costs and rewards, and $p(A|I)$ is the probability that the agent would take the observable actions A given the selected intention I. Eq. (7) shows how the Naïve Utility Calculus captures two types of rationality. $p(I|C, R)$ captures the expectation for rational choice (selecting an intention that maximizes utilities; Eq. (5)), and $p(A|I)$ captures the expectation for rational action (taking actions so that the intention minimizes costs and thus produces the expected utilities; Eq. (4)).

In our implementation, we solve Eq. (6) through Monte Carlo samples of cost and reward functions. Our application of softmax at a choice and at an action level enables us to infer the cost and reward functions with fewer samples. Under a strict assumption of utility maximization, only cost and reward functions where the agents' actions yield the highest utility produce positive probabilities. Consequently, most samples have probability 0. By softmaxing utility maximization, cost and reward functions that are far away from the true underlying costs and rewards have near zero probability, but this probability increases as the sampled functions get closer to values under which the agent has truly maximized utilities. Thus, through softmax, cost and reward functions under which the actions are reasonable, even if they are not optimal, reveal the agent's underlying costs and rewards.

As in previous Bayesian models of human action understanding, but even more so for the more complex planning-based inferences our model considers, we do not expect that human observers will be exactly or explicitly implementing the computations specified in Eqs. (6) and (7). For instance, it would be sufficient both for everyday action understanding, and for the purposes of our modeling here, if people were able to approximate the likelihood (Eq. (7)) with one or a few high-probability plans rather than a sum over all possible plans. In our model, we approximate the posterior distribution through Monte Carlo likelihood weighting: we first sample n cost and reward functions from the prior distribution and compute the likelihood that each sample produces the observed actions. Under this approach, the prior information is encoded in the frequency and distribution of the samples, and the likelihood is integrated as a weighting factor.

2.2.1. Qualitative model predictions

Fig. 3 shows example trials that highlight the types of inferences that our model aims to capture. In the two events in Fig. 3a (tested in Experiment 1a), the astronaut takes a longer route to get to the space station. In the first event, the astronaut's behavior is ambiguous and can be explained by setting a high reward on the orange care package, a low cost on the pink terrain, or both. In the second event, however, the observed behavior can only be explained by utility functions where the cost to traverse the purple terrain is lower than the pink terrain and where the orange care package has no reward. Fig. 3b shows two events (tested in Experiment 1b) where the astronaut chooses to walk into one of two terrains to collect a care package. In the first event, many utility functions can explain the astronaut's behavior, as the distance to the care package was lower, predicting that the event should be more ambiguous

for observers. By contrast, utility functions that can explain the second event must have a high reward associated with the white care package (relative to the orange care package), a low cost associated with the purple terrain (relative to the pink terrain), or both. Finally, Fig. 3c shows two events (tested in Experiment 1c) where the agent can obtain a care package through a minimal detour. Here, the first event is consistent with multiple utility functions: the astronaut might not like the care package, or she might find the pink terrain too difficult to walk through. By contrast, the behavior in the second event can only be explained by a utility function where the care package is rewarding, and the pink terrain is costly relative to the yellow one.

Note that while deriving the predictions of the Naïve Utility Calculus in events like the ones we describe above is simple, all of these inferences were qualitative and are consistent with an infinite number of different cost and reward functions that satisfy those properties. By formalizing the Naïve Utility Calculus as a computational model, we can make quantitative predictions about the precise shape of each cost and reward function that we test quantitatively in our studies.

2.3. Relation to reinforcement learning frameworks

Research in machine learning and AI has approached similar problems, but our work differs in a few critical ways. Conceptually, our generative model can be thought of as a form of model-based hierarchical reinforcement learning (MB-HRL; Sutton, Precup, & Singh, 1999; Botvinick & Weinstein, 2014), an extension of reinforcement learning where policies are expressed in terms of *options*—sub-policies that complete different tasks (e.g., options may allow an agent to plan in terms of reaching key locations in the environment). Our model has a similar hierarchical structure, where high-level plans are expressed in terms of goals. However, one key difference between our generative model and HRL models is that we use goals to determine which components of the reward function to pursue and which to forego (capturing the intuitive difference between *liking* something, and actively *wanting* it). By contrast, options in HRL usually do not affect the reward function and instead only help agents plan over extended spatio-temporal scales.

Another difference between our generative model and RL models more generally is that we distinguish between costs and rewards (rather than using a single notion of rewards that can be positive or negative). We do this because, in our model, agents can select which rewards to pursue and which to ignore, but they are always subject to the costs of their actions. As a consequence, our model separates costs and rewards at different levels of the hierarchy (see Fig. 2): Goals are built by attending to the reward function, while each goal's policy is built by attending to the cost function.

Finally, inference over our generative model is conceptually similar to *inverse reinforcement learning* (IRL; Ng & Russell, 2000; Ramírez & Geffner, 2011; Ziebart, Maas, Bagnell, & Dey, 2008) and *inverse optimal control* (Dvijotham & Todorov, 2010; Mombaur, Truong, & Laumond, 2010), where the goal is to infer the reward function behind observed optimal behavior. Our work differs from these AI approaches in that our goal is to explain human social reasoning, which requires modeling human-like inferences over a generative model that aligns with our intuitive theory of how other people act. Our model therefore treats observable actions as traces from an approximate planner (instead of an optimal policy), and we use inference procedures based on how people attribute unobservable goals (Baker et al., 2009; see Jara-Ettinger, 2019 for a review of Inverse Reinforcement Learning in the context of Theory of Mind research).

2.4. Alternative models

As reviewed above, empirical work on action-understanding qualitatively fits the predictions of the Naïve Utility Calculus (see Jara-Ettinger et al., 2016 for review). Nonetheless, people might arrive at these inferences in many ways, only some of which would be best characterized as implementations of or approximations to the formal computations described above. Here we sketch two broad classes of alternative accounts: variations of the Naïve Utility Calculus and a set of domain-specific heuristics. The goal of our experiments is thus not only to find evidence for our account but also to distinguish it from other (often simpler) potential explanations that can generally get things right. These heuristics and model lesions are described in more detail in the sections where they become relevant, but for the sake of completeness we summarize them here.

2.4.1. Naïve Utility Calculus variations

The first class of alternative models posits that people have a Naïve Utility Calculus that is different or more limited than the one we have presented. We specifically consider three alternative hypotheses.

The first alternative hypothesis is that people expect agents to maximize the rate of rewards they obtain, rather than the difference between costs and rewards. This model is inspired by research suggesting that, in some contexts, humans maximize the rate of rewards that they obtain, rather than the difference between costs and rewards (Bettinger & Grote, 2016; Constantino & Daw, 2015). Indeed, all qualitative data reviewed in the introduction is consistent with a rate-maximization model, and it is thus possible that action understanding is also structured around this expectation. The rate-based model is identical to our main model with the exception that the utilities are defined as the ratio of rewards to costs, rather than the difference:

$$U(a, o) = R(a, o)/C(a, o) \tag{8}$$

This model produces similar qualitative predictions to our main model, but it makes different quantitative predictions in all experiments, and we extensively evaluate it in Experiment 1.

Sticking to the more conventional difference formulation, our second alternative hypothesis is that people use a Naïve Utility Calculus to infer costs and rewards, but that, for the sake of cognitive efficiency, they rely on the minimal number of events necessary

to draw these inferences, rather than considering every observed action in every event. More concretely, the full Naïve Utility Calculus model integrates all observations to infer the costs and rewards that best explain all of these observations. For instance, when we watch an agent act in two different situations, Eq. (6) becomes

$$p(C, R|A_1, A_2) \propto l(A_1, A_2|C, R)p(C, R) \tag{9}$$

In practice, however, integrating information from multiple action sequences may be more work than necessary: if an agent's costs and rewards are constant across time, it would be easier to infer them from a single event. For instance, using only the first observation or the most recent observation may be sufficiently accurate, even if this leads to suboptimal inferences in some cases. Such a model would make identical predictions to the full model in most of our tasks, but it diverges when observers can watch the agent over repeated events or interactions and we test this possibility in Experiment 2.

Finally, our third alternative hypothesis is that basic goal attribution is not generally sensitive to the cost-benefit tradeoffs underlying the choice. We formalize this hypothesis through a model of the most basic Bayesian inverse-planning approach to goal inference in Baker et al. (2009). While this model cannot make the kinds of inferences that we test in most of our experiments, it can be used for predicting an agent's future goals. We thus evaluate it in Experiment 4.

2.4.2. Domain-specific heuristics

The second kind of alternative that we explore is one where people rely on simple heuristics that together serve as a reasonable approximation to the Naïve Utility Calculus in some common circumstances. To our knowledge, there is no current framework that posits a collection of heuristics which, together, approximate the types of social inferences that we focus on here. The heuristics we present here thus do not reflect an overarching framework for approximating Bayesian computation in all domains, and were instead designed to exploit natural features of observable behavior that provide direct evidence of the underlying costs and rewards. None of these accounts is sufficiently broad to make predictions in all four of our experiments. Nonetheless, each captures the broad qualitative structure of the judgments in at least one of our tasks.

One possibility is that participants infer an agent's costs and rewards by tracking the amount of time an agent spends in each terrain and which objects she collects. We test a heuristic of this kind in Experiments 1 and 3. When predicting future goals, participants might generalize only on the basis of the terrain the agent prefers, or on the basis of the object that the agent prefers. We test this heuristic in Experiment 4. When inferring whether an agent knew her own costs and rewards, participants may simply associate changes of behavior with ignorance. We test this heuristic in Experiment 5. Finally, when deciding whether someone is nice or mean, people may judge niceness solely based on whether the agent helps. We test this heuristic in Experiment 6. We discuss these heuristics in more detail in the experiments where they come into play. To preview our results, we find that heuristics of this sort can often account for much of the variance in our participants' judgments, but are too coarse. People consistently make fine-graded inferences that our Naïve Utility Calculus model explains, but which these various simple heuristics fail to capture.

3. Experiment 1

We begin by testing the main prediction of the Naïve Utility Calculus: people should be able to jointly infer agents' unobservable costs and rewards by assuming that agents maximize utilities. In Experiment 1 participants watched an agent navigate a world with different kinds of terrains and care packages and they were asked to infer the cost and reward functions. To ensure that our results do not hinge on any specific type of behavior or spatial layout, we used three different geometries across Experiments 1a-1c. In Experiment 1a (Fig. 4) we considered situations where the shortest paths might involve crossing multiple types of terrains, and where detours could be explained either by appealing to differences in costs or differences in rewards. In Experiment 1b (Fig. 5) we considered situations where the astronauts were either forced to cross one type of terrain, or where they could choose whether to venture into a terrain to collect a care package. Finally, in Experiment 1c (Fig. 6) we considered situations where agents could take detours on their way to the space station that were physically minimal, but nonetheless revealed agents' costs and rewards.

Throughout, we evaluate participants' responses against our Naïve Utility Calculus model and two alternative models. The first alternative model is the rate-based model, which defines utilities as the rate of rewards over unit of cost instead of their difference. The second alternative model is a heuristic that maps the length of the path in each terrain onto cost judgments, and directly maps the choices of the collected care packages onto reward judgments. All experiments were approved by MIT's IRB under protocol #0812003014 ('Learning and Reasoning with Words and Concepts').

3.1. Methods

Participants. 90 participants (mean age: 31.96 years), range 21–60 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform (n = 30 per experiment).

Stimuli. Figs. 4–6 show the stimuli. Each trial consisted of a map with one or two terrains, one or two care packages, a starting and an ending state, and the path an agent took from the starting to the end state. Experiment 1b used an additional terrain in the center that was treated as neutral and participants did not have to infer its cost. The stimuli were built combinatorically, using three different basic geometries, one per experiment (see Appendix for details). Experiment 1a consisted of 16 trials, Experiment 1b of 17 trials, and Experiment 1c consisted of 14 trials. In each display we used three possible terrain textures and two object drawings to generate up to twelve different versions of each trial (depending on number of terrains and objects in stimulus; see Fig. 3 for examples and Appendix for full space of stimuli) and participants saw a random one in each trial.

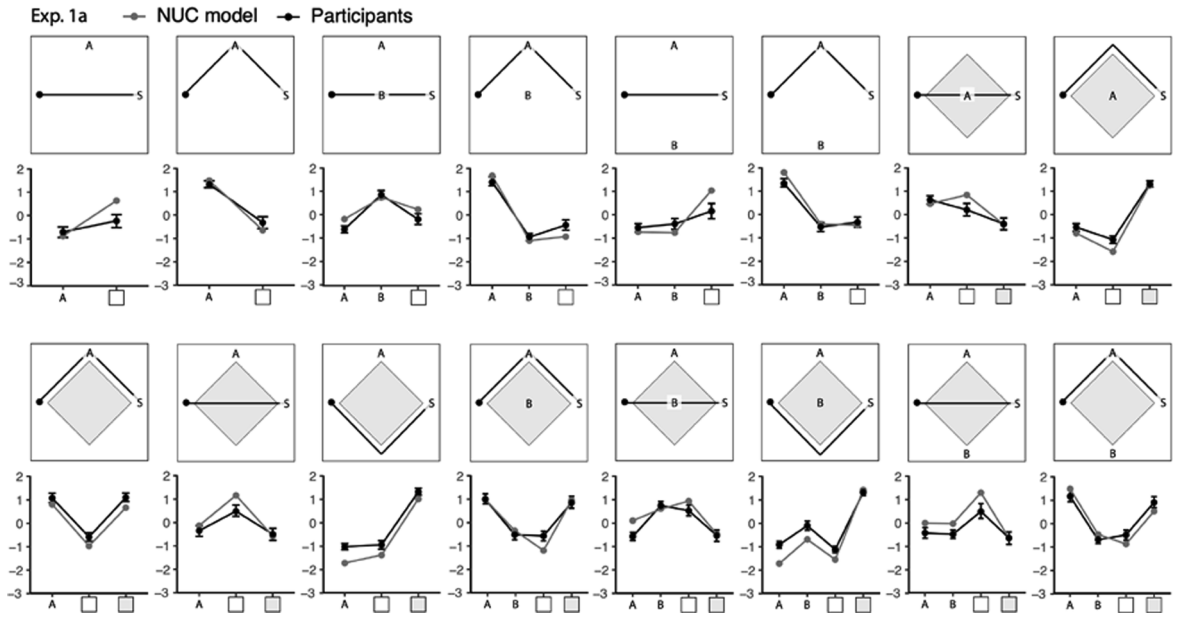

Fig. 4. Stimuli schematic (see Fig. 3 for examples of actual stimuli) and results for Experiment 1a. Each trial's results appear below their corresponding stimuli. Gray curves show the model's inferred function (z-scored within judgment type) and black curves show average participant judgments (z-scored within judgment type for each participant and averaged). Higher values on the reward dimensions indicate higher rewards and higher values on the cost dimensions indicate higher costs. Horizontal black bars are 95% confidence intervals.

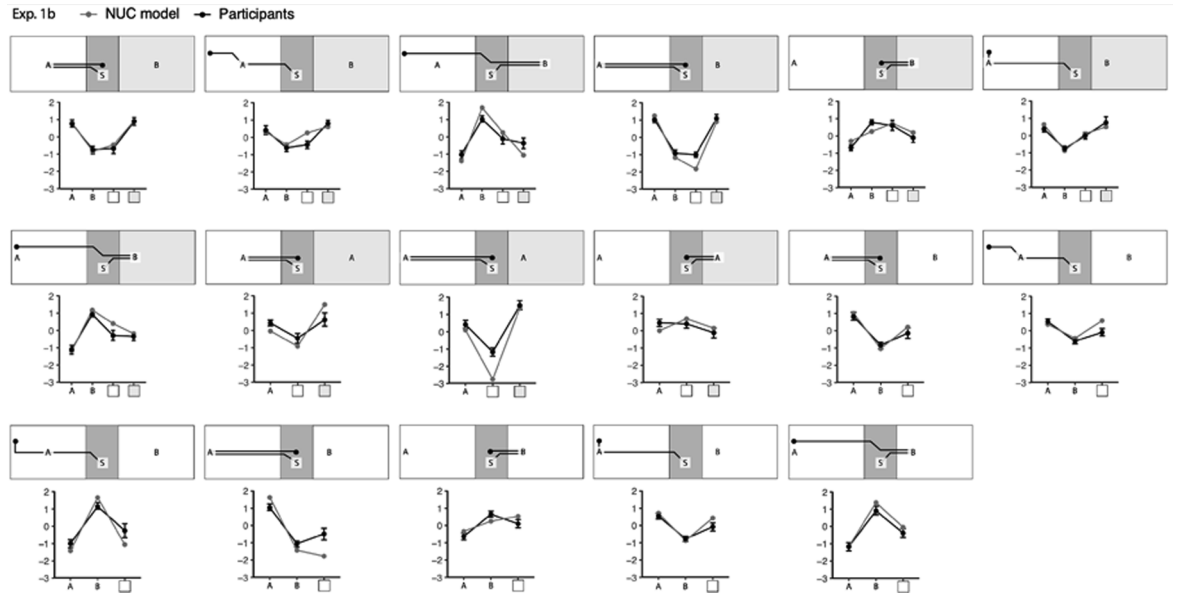

Fig. 5. Stimuli schematic (see Fig. 3 for examples of actual stimuli) and results for Experiment 1b. Each trial's results appear below of their corresponding stimuli. Gray curves show the model's inferred function (z-scored within judgment type) and black curves show average participant judgments (z-scored within judgment type for each participant and averaged). Higher values on the reward dimensions indicate higher rewards and higher values on the cost dimensions indicate higher costs. Horizontal black bars are 95% confidence intervals.

Procedure. Participants first read a cover story that explained the task. In the story, participants learned that they would watch astronauts land in an alien planet with novel terrains and travel towards the space station. Participants were told that the astronauts could always travel all types of terrain and that they could collect up to one care package on their way home. Participants were told their task was to determine the astronauts' ability to travel each kind of terrain (costs), and their desire to collect each care package (rewards). Participants then completed a multiple-choice questionnaire designed to ensure they understood the task. Specifically, participants were asked (1) if the astronauts have identical or different abilities (different astronauts have different abilities); (2) if the astronauts had different or identical care package preferences (different astronauts have different preferences); (3) if astronauts

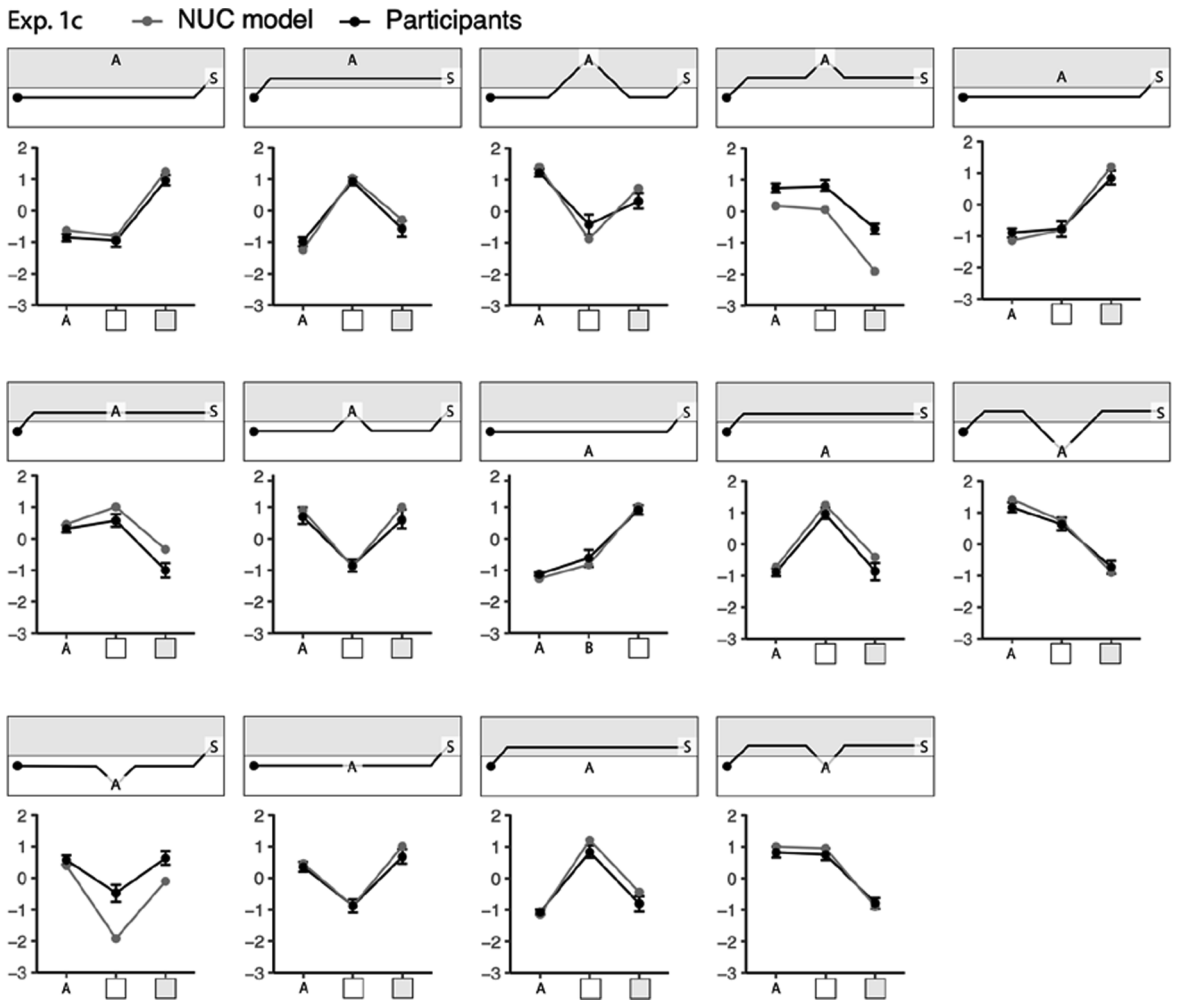

Fig. 6. Stimuli schematic (see Fig. 3 for examples of actual stimuli) and results for Experiment 1c. Each trial's results appear below of their corresponding stimuli. Gray curves show the model's inferred function (z-scored within judgment type) and black curves show average participant judgments (z-scored within judgment type for each participant and averaged). Higher values on the reward dimensions indicate higher rewards and higher values on the cost dimensions indicate higher costs. Horizontal black bars are 95% confidence intervals.

could always cross all terrains, even if it is exhausting (all astronauts can travel across all terrains); and (4) whether astronauts had to collect at least one care package, up to one care package, or all care packages (in this experiment, the astronauts could collect up to one care package). Participants were given access to the experiment only if they responded all questions correctly. They were otherwise redirected to the beginning of the tutorial and asked to re-read the cover story if they wished to participate.

In each trial, participants saw a path on the left side of the screen, and up to four questions on the right side (the number of questions depended on how many types of terrains and how many types of care packages appeared in the stimulus). The cost questions asked "How easy is it to cross this terrain?" followed by a small picture of the relevant terrain. The reward questions asked "How much does he like this container?" with a picture of the care package. The responses were input using a discrete scale ranging from 0 to 10 with accompanying text on values 0, 5, and 10. In the cost (for crossing terrain) questions, 0 indicated "extremely easy", 5 indicated "average", and 10 indicated "extremely exhausting." In the reward (for collecting care packages) questions, 0 indicated "not at all", 5 indicated "average", and 10 indicated "a lot."

3.2. Results

We z-scored model predictions and participant judgments to match their scales. Model predictions were split into cost and reward judgments and z-scored within each category (such that the average cost prediction and the average reward predictions are 0). Participant cost and reward judgments were z-scored separately for each subject (such that for each participant, their average cost inference and their average reward inference is 0) and they were then averaged across participants. Figs. 4–6 show each trial's results and Fig. 7 shows participant's overall inferred cost and reward functions along with model predictions. Overall, participant

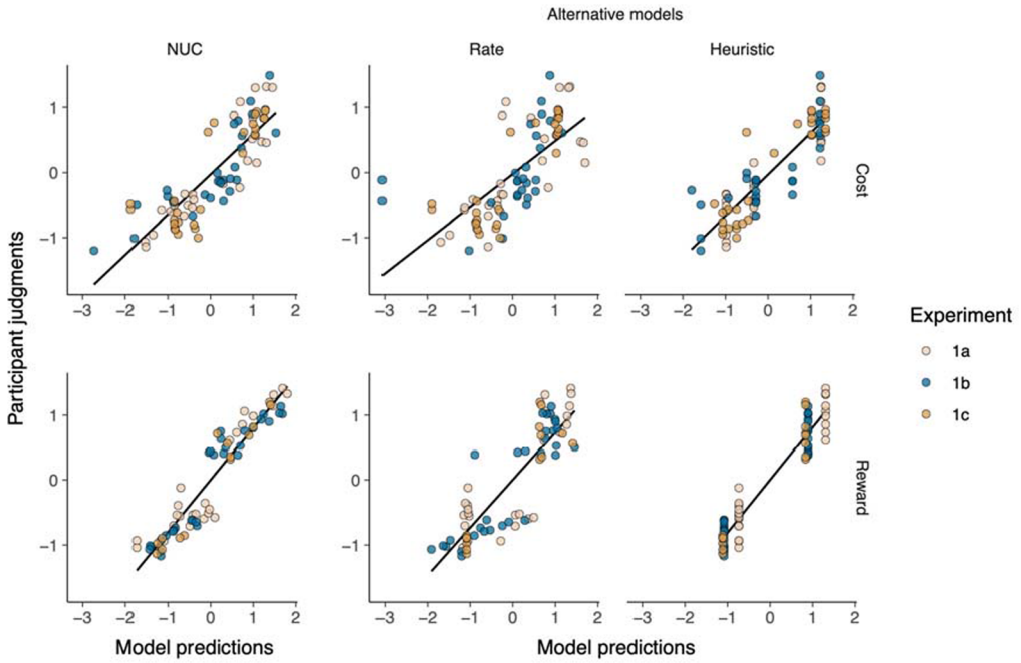

Fig. 7. Results from Experiment 1. (a) Each subplot shows a comparison between a model and a question type (costs or rewards). Each point represents a judgment. The x-axis shows the model predictions (z-scored) and the y-axis participant judgments (z-scored within participant and averaged across participants). In each point, the x-axis corresponds to the model prediction and y axis to the participant judgments.

judgments closely tracked the quantitative predictions of our model with a 0.85 correlation (95% CI: 0.81–0.90) on cost inferences and a 0.95 correlation (95% CI: 0.93–0.97) on reward inferences, producing a global correlation of 0.90 (95% CI: 0.93–0.07).

Having established that the Naïve Utility Calculus (NUC) predicts participant judgments with quantitative accuracy, we tested if the definition of utilities as the difference between rewards and cost fits participant judgments better than an alternative formulation where utilities are defined by the reward rate (see Alternative models section). The rate-based model was identical to our main model with the differences that it computed utilities as $U(a,o) = R(a,o)/C(a,o)$ rather than $U(a,o) = R(a,o)\text{-}C(a,o)$. Both models make the same qualitative predictions, but quantitatively, the rate model did not track participant judgments as well, with a correlation of 0.70 (95% CI: 0.59–0.80) on cost inferences and a correlation of 0.87 (95% CI: 0.82–0.92) on reward inferences, respectively (see Fig. 7). The correlation difference between the NUC model and the rate model was reliably above zero for both costs (95% CI of NUC correlation minus rate-based correlation: 0.06–0.25) and rewards (95% CI of NUC correlation minus rate-based correlation: 0.03–0.12).

Finally, we tested if participants relied on simple heuristics that approximate the responses of our NUC model. Specifically, we tested if participants estimate the costs and rewards directly from the amount of time the agent spends on each terrain, and on which objects the agent collects. To test this possibility, we computed costs as the total distance that the spaceman traveled in each terrain (using a unit of 1 per square), and rewards as 1 or 0, based on whether the agent collected the care package or not. Having calculated these costs and rewards for each trial, we then z-scored cost judgments and reward judgments separately, in the same way we did for our NUC model. The heuristic model correlated highly with participants, with a 0.88 (95% CI: 0.84–0.93) correlation on cost judgments and a 0.96 (95% CI: 0.95–0.97) correlation on reward judgments. However, as Fig. 7 suggests, this correlation was high only because it captured the coarse structure of participant judgments and not necessarily the fine-grained structure. The strongest prediction our heuristic makes is that, in some cases, different stimuli should produce identical inferences in costs or rewards (either because the astronaut collects and ignores the same care packages, or because the astronaut spends the same amount of time in each terrain in both stimuli), visible as vertical columns of dots in Fig. 7. If the heuristic is correct, then any graded variability within each predicted cluster of judgments should be the result of noise in participant judgments. Consequently, judgments within each cluster should not correlate with inferences from our Naïve Utility Calculus model. We tested this possibility by splitting our results into clusters where the heuristic predicted that participants should produce the same judgment, and we correlated these sub-regions against the Naïve Utility Calculus. Because several of the regions had too few data points for correlation to be interpretable, we only considered the twelve regions that had at least six data points (six cost regions and six reward regions; five regions using data from Exp 1a, four regions using data from Exp 1b, and three regions using data from Exp 1c). All twelve sub-regions positively correlated with the predictions of the Naïve Utility Calculus (mean correlation = 0.65; see appendix for full details), showing that the variability that the heuristic fails to capture stems from participant sensitivity to features of the stimuli that the Naïve Utility Calculus captures but that the simple heuristic does not.

Together, these results show that people can break down observable behavior into judgments of an agent's unobservable costs and

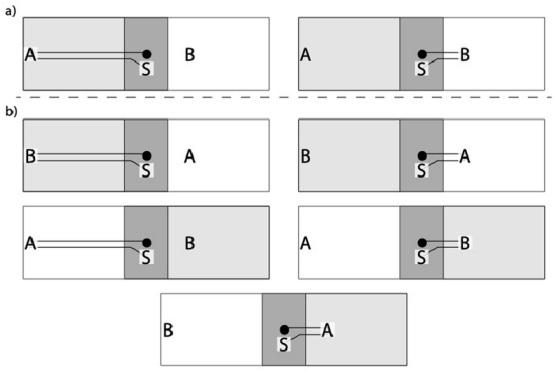

Fig. 8. Stimuli from Experiment 2. Each trial consisted of two events. Five trials consisted of the top left event in panel a, and one of the five events on panel b. The other five trials consisted of the top right event in panel a, and one of the five events on panel b. For the experiment, we generated up to twelve different version of each stimulus (by varying the colors of the terrain and colors of the care packages; see Appendix for different versions of each stimulus) and participants were shown a random one for each of the sixteen trials.

rewards; that these inferences are consistent with an expectation of utility maximization; and that, although simple heuristics approximate these inferences, participant answers show more sensitivity than a simple heuristic predicts.

4. Experiment 2

Experiment 1 looked at judgments from a single event. However, in the real world, we often observe people acting on multiple occasions, frequently under different contexts, and subsequent observations may lead us to revise our beliefs about what they want and what they can do. In Experiment 2 we test how participants infer agents' underlying costs and rewards when there are multiple action events. This experiment serves two purposes: first, to test if people's inferences over repeated events are predicted by the generative model of the Naïve Utility Calculus, and second, to test if participants' inferences might also be well explained by a more limited Naïve Utility Calculus model, where, for computational efficiency, inferences are performed over just a single event (either the first or the most recent observation).

4.1. Methods

Participants. 30 participants (mean age (SD): 35.31 years (11.86 years), range 18–62 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform.

Stimuli. Each stimulus consisted of two maps visually similar to the ones used in Experiment 1b. The stimuli were built by creating the full space of possible stimuli and removing stimulus that were (1) irrational under any cost-reward decomposition or (2) where the second map provided no information that the first map already did. Fig. 8 shows the final ten stimuli. For each trial, we generated twelve different versions of the same stimuli varying the colors of the terrains and care packages and participants were shown a random one for each of the ten trials.

Procedure. The procedure was identical to Experiment 1 with the exception that on each trial participants saw two stimuli of the same agent traveling over two different landscapes in two different situations, with different terrains and package locations in each situation.

4.2. Results

Model predictions and participant judgments were z-scored in the same way as in Experiments 1. Fig. 9 shows model predictions (x-axis) plotted against participant average judgments (y-axis) for our main model and our two alternative models. Participant judgments showed a high correlation with the full model which combined several observations in a Bayesian way (see Eq. (9)). Our model showed a correlation of 0.86 (95% CI: 0.78–1) on costs and a correlation of 0.92 (95% CI: 0.86–1) on rewards.

It is possible that participants did not integrate information from both events, and that, instead, each event contained enough information to recover the underlying costs and rewards. To test this, we compared participant judgments against two reduced models. One model only relied on the first event to infer costs and rewards (First only reduced model) and the second model only relied on the last event to infer costs and rewards (Second only reduced model). The right side of Fig. 9 shows the alternative model predictions against participant judgments. Predictions relying only on the first observed event showed reliably lower correlations: 0.61 (0.25 lower than the main model; 95% CI: 0.08–0.4) and 0.69 (0.23 lower than main model; 95% CI: 0.02–0.37) for costs and rewards respectively. Predictions relying only on the last observed event also showed reliably lower correlations: 0.57 (0.29 lower than the main model; 95% CI: 0.09–0.46) and 0.68 (0.23 lower than main model; 95% CI: 0.05–0.37) for costs and rewards respectively. These results thus suggest that, when people infer the underlying costs and rewards that explain another person's behavior, they can integrate information from multiple events to make more precise cost and reward inferences.

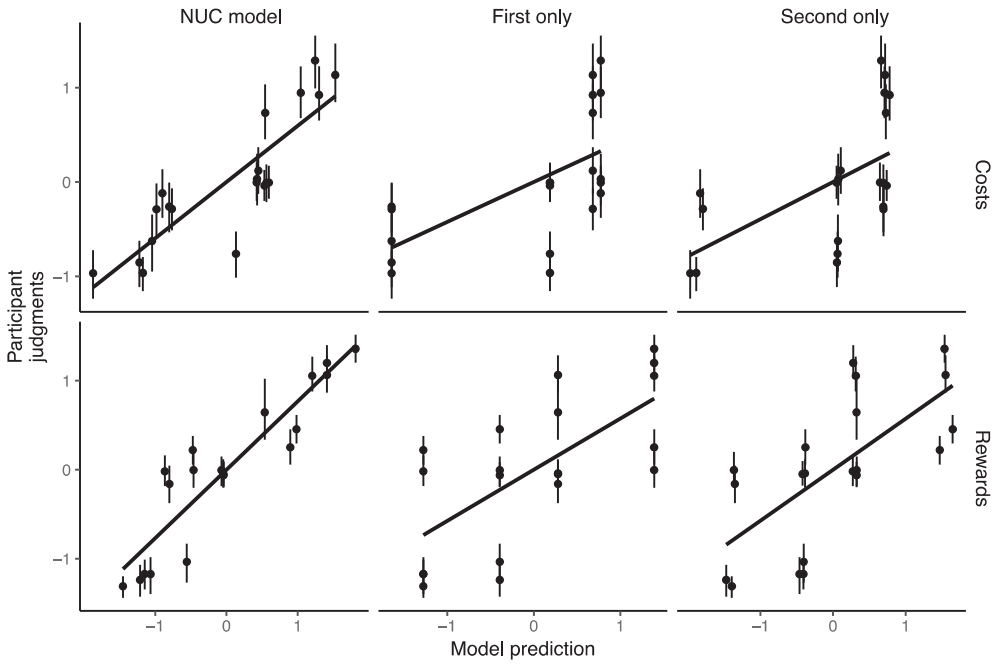

Fig. 9. Results from Experiment 2. Each dot represents a trial where the x-axis shows the model prediction and the y-axis shows average participant judgments. The pair of plots shows the model's inferences when two trials are combined. The reduced models show the model fits when only the first event is used to infer costs and rewards (first only) and when only the last event is used (second only).

5. Experiment 3

Experiments 1–2 show that people can quantitatively decompose agents' behavior into judgments about their costs and rewards, relying on the assumption that agents maximize utilities. In Experiment 3 we compare people's relative confidence judgments over cost and reward inferences against the confidence of our generative model. That is, rather than asking participants for point estimates (as in Experiments 1–2), we asked them to determine their relative confidence over which object had a higher reward, and which terrain had a higher cost.

Confidence judgments about cost and reward inferences also correlate with superficial features of the stimuli. If an agent walks a long path in one terrain, rather than walking a short path in another terrain, then, intuitively, we can be confident that the first terrain is easier. By contrast, if an agent walks a short path in one terrain, rather than walking a much longer one in another terrain, then, our confidence decreases. Thus, in this experiment we compare participant judgments to our Naïve Utility Calculus model as well as to a simple heuristic that estimates confidence based on the relative distance of alternative plans.

5.1. Methods

Participants. 30 participants (mean age (SD): 37.97 years (9.08 years), range 22–55 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform.

Stimuli. The stimuli were the same used in Experiment 1b. See Fig. 4.

Procedure. The procedure was identical to Experiment 1 with the exception that participants were presented with two slidebars initialized in the middle. The first question asked "Which container does the astronaut like best?" Each end of the slidebar said "Definitely" accompanied by a picture of the care package, and the center said "Not sure." The second question asked "Which terrain is easier to cross?" with each edge saying "Definitely" accompanied by a picture of the terrain, and the center saying "Not sure." Some trials did not have two rewards and some did not have two terrains (see Fig. 4 for stimuli). In those cases, only the questions where there were two alternatives were asked.

5.2. Results

Model predictions and participant judgments were z-scored in the same way as in Experiment 1. Fig. 10 shows participant's judgments on each trial along with model predictions. Overall, participants showed a correlation of 0.96 (95% CI: 0.94–0.99) with our model. We next compared if these judgments could stem from a simpler heuristic. Building on the logic of the heuristic evaluated in Experiment 1, the distance spent on a certain terrain, or pursuing an object may directly guide people's confidence judgments. To test this alternative, we designed a simple heuristic that estimates confidence by considering the relative distance the agent incurred to get one object.

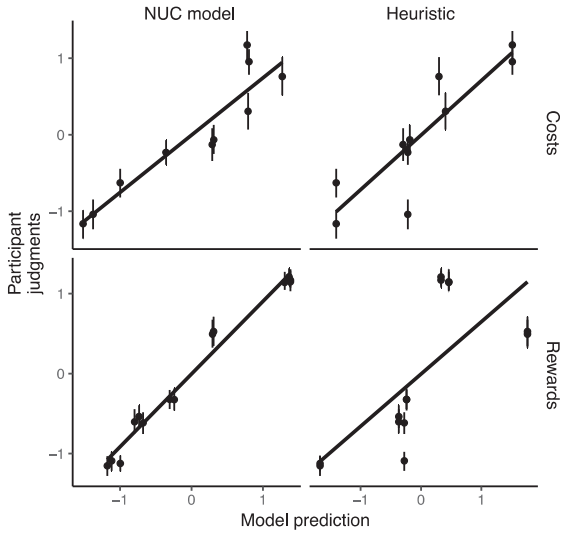

Fig. 10. Results from Experiment 3. Each dot represents a trial. The x-axis shows the model predictions (z-scored) and the y-axis shows participant judgments (z-scored). Vertical lines show 95% confidence intervals.

In this heuristic, we estimated confidence over rewards and costs by dividing the number of actions the agent took to her goal, by the number of actions the agent would have needed to take to complete the alternative goal (note that in this experiment any trial had at most two kinds of objects). For instance, if an agent took 30 actions to obtain an object, when she could have obtained the competing object with only 3 actions, then the raw confidence judgment would be 10 (high confidence). By contrast, if an agent took 3 actions to obtain an object when she could have obtained the competing object with 30 actions, then the confidence is 0.1 (low confidence). We used this as a measure of confidence over costs as well as rewards. The heuristic predictions were z-scored.

Fig. 10 shows participant inferences (y-axis) against the NUC model and the heuristic (x-axis), split by cost and reward inferences. Our heuristic captured the rough qualitative structure of people's judgments, showing a correlation of 0.77 (95% CI: 0.67–0.90) with participant judgments. Nonetheless, the Naïve Utility Calculus showed a reliably high correlation relative to the heuristic (difference = 0.19; 95% CI: 0.06–0.31). These results expand on the results from Experiment 1, showing that, beyond predicting the cost and reward functions that people infer from observable actions, the Naïve Utility Calculus model also predicts people's relative confidence in these inferences.

6. Experiment 4

In Experiment 4 we test if people can use inferred costs and rewards to predict how an agent will behave in a new situation. We then compare participants' predictions against our full model, against a simple goal-inference model, and against two plausible heuristics that approximate our model predictions using simple queues.

Our goal-inference model is an implementation of the model in Baker et al., 2009. This model can be thought of as a simplified model of the Naïve Utility Calculus where plans always consist of a single goal, and where costs are fixed, observable, and do not vary across agents or terrains. Nonetheless, this goal-attribution model is sensitive to relative distances, and infers stronger goals as distances increase. This model could not be used as an alternative model in Experiment 1 because it does not make inferences beyond goal-attribution, but it can be used to predict future goals. Because our main model predicts behavior by relying on the inferred cost and reward functions, this alternative model enables us to test if participants are indeed predicting future goals through their Naïve Utility Calculus, or whether they are simply inferring the agent's current goal, and using this to predict the agent's next goal.

Our two heuristics consist of simple mappings from features of the stimuli onto future goals. The first heuristic (object-tracking heuristic) that we compare participant judgments against assumes that the agent will continue to go towards the object that she showed a preference for in the past. The second heuristic (terrain-tracking heuristic) assumes that the agent will continue to choose the terrain that she navigated through in the past.

6.1. Methods

Participants. 30 participants (mean age (SD): 31.33 years (8.73 years), range 22–54 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform.

Stimuli. Participants were presented with stimuli each consisting of two maps: one map with the astronaut's path, and a second map with the astronaut in a starting location and no displayed path. Nineteen stimuli were used in total (see Fig. 11; see Appendix for details). The stimuli were built by creating the full space of possible stimuli and removing stimulus where (1) the prediction world was the same as the observed world (2) where the observed trial gave full or no information about the astronaut would do in the new

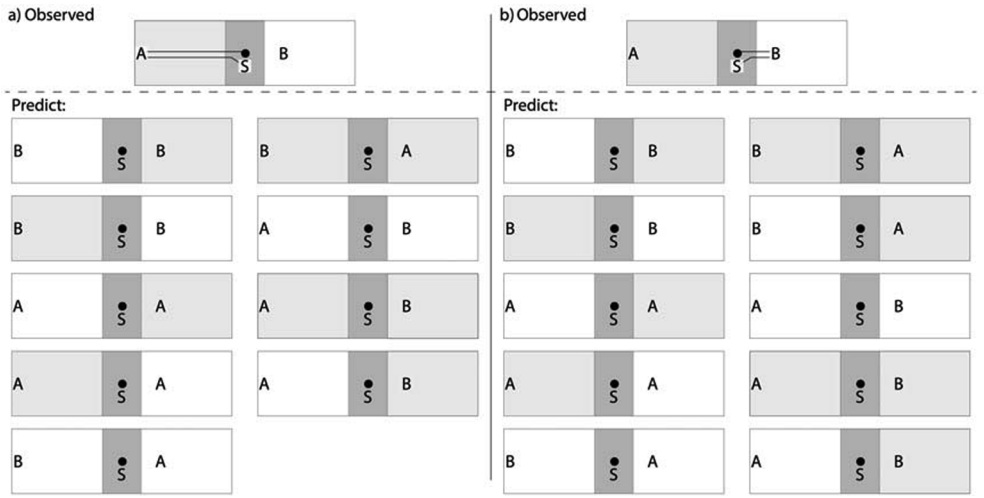

Fig. 11. Schematic of stimuli used in Experiment 4. See Fig. 3d for picture of example stimulus. (a) Trials where participants saw the agent choose the goal on the left and then had to predict their behavior in the 9 maps that appear below it. (b) Trials where participants watched the agent choose the goal on the right and were then asked to predict what the agent would do in the 10 maps that appear below it.

world (see Appendix for full list and detailed descriptions). For each trial, we generated twelve different versions of the same stimuli varying the colors of the terrains and care packages (see Appendix for full space of stimuli) and participants were shown a random one in each of the nineteen trials.

Procedure. The procedure was similar to Experiment 1. However, instead of judging the astronaut's costs and rewards, participants were asked to predict where the astronaut would go in a new map by using a slidebar that ranged from 0 ("definitely left") to 1 ("definitely right") with the text "not sure" marked at 0.5.

6.2. Results

Model predictions and participant judgments were z-scored in a similar way to Experiment 1 (see Results section of Experiment 1), with the difference that responses were not split into cost and reward judgments (because here participants only predicted the future goal). The top left panel of Fig. 12 shows people's judgments compared to the full Naïve Utility Calculus (NUC) model predictions. Participants' predictions correlated highly with the NUC model ($r = 0.87$; 95% CI: 0.81–0.93), suggesting that people are not just able to decompose other people's behavior into judgments of costs and rewards, but to later recombine these judgments to predict future goals.

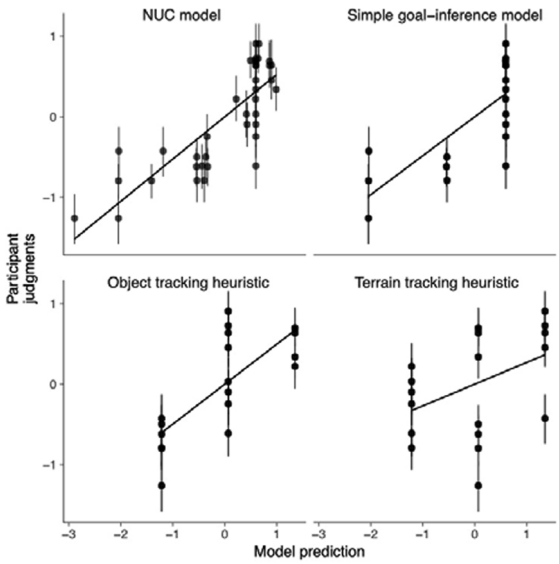

Fig. 12. Results from Experiment 4. Each dot represents an experiment trial. The x-axis shows the model prediction and the y-axis shows participant judgments. Horizontal lines represent 95% confidence intervals and diagonal lines show best linear fit.

The top right panel of Fig. 12 shows participant judgments compared to the predictions of the simple goal-attribution model. Participant judgments showed a correlation of 0.74 (95% CI: 0.59–0.97) with the goal-inference model. This model predicts that participants should only give three kinds of answers: full confidence that the agent will go for the left goal, full confidence that the agent will go for the right goal, and complete uncertainty. If the model is correct, then any deviation from these three kinds of judgments should be the result of experimental noise. To evaluate this possibility, we used the same analyses from Experiment 1: We computed the correlation between the Naïve Utility Calculus (NUC) and the simple goal-inference model's predicted regions of indifference. If this variability is noise, then it should not correlate with NUC model predictions. Only the rightmost region contained enough data (left and middle regions each had n = 3 data points) for this analysis and participant judgments in this region showed a correlation of $r = 0.81$ with our NUC model, showing that participants judgments were indeed driven by cost and reward inferences rather than by simple goal-attribution.

To evaluate whether people's judgments resulted from a simpler heuristic, we considered two alternative models. The object tracking heuristic predicts that the agent will get the same kind of care package that she collected in the original trial and that she will select one at random when both sides have the same kind of care package. The terrain tracking heuristic predicts that the agent will get the care package in the terrain that she has already travelled and that she will select one at random when both sides have the same terrain. Fig. 12 (bottom panels) shows the heuristic predictions (x axis) against participant judgments (y axis). The object tracking heuristic showed a correlation of r = 0.76 (95% CI: 0.63–0.95) while the terrain tracking heuristic showed a substantially lower correlation of r = 0.42 (95% CI: 0.12 – 0.78).

Because the terrain tracking heuristic did not explain the qualitative pattern of data, we did not evaluate it further. Because the object tracking heuristic fit the qualitative pattern of data, we tested if the variability in participant judgments in regions where the heuristic predicts no difference correlated with NUC model predictions. The leftmost heuristic region (z-scored prediction = -1.21; bottom left plot in Fig. 12) marginally correlated with the Naïve Utility Calculus (r = 0.78; p = 0.07; n = 6 data points), the middle heuristic region (z-scored predictions = 0.07; bottom left plot in Fig. 12) significantly correlated with the Naïve Utility Calculus (r = 0.85; p = 0.007; n = 8 data points), and the right heuristic region (z-scored predictions = 1.35 bottom left plot in Fig. 12) showed a positive correlation (r = 0.36) although this was not significant (p = 0.56; n = 5 data points). Together, the lower heuristic fits relative to the NUC model, and the finding that the NUC correlates in regions that the heuristic predicts should be random noise shows that participants did not rely on this object tracking heuristic. Instead, our results suggest that, when people predict future actions, they estimate the agent's utility function in the new environment, and assume that the agent will maximize this function.

7. Experiment 5

Experiments 1 and 2 suggest that people have a Naïve Utility Calculus that enables them to infer other people's costs and rewards and predict future events based on these inferences. In these situations, however, the agent herself always knew the costs and rewards. In more realistic circumstances, agents can be either naïve or wrong about the costs and rewards involved (e.g., Jara-Ettinger et al., 2016; Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017; Moutoussis, Dolan, & Dayan, 2016), act based on impartial information, and update their beliefs as they obtain new experience. In Experiment 5 we test if people can infer whether an agent is knowledgeable or ignorant given how she does or does not revise her behavior across situations. We compare participant judgments against our full model, and against a simple heuristic that assumes that agents are naïve whenever they revise their behavior, and that they are knowledgeable when they do not.

7.1. Methods

Participants. 30 participants (mean age (SD): 34.17 years (8.00 years), range 21–52 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform.

Stimuli. Schematics of the stimuli are shown in Fig. 13 (see Fig. 3e). Each stimulus consisted of two maps showing the astronaut's trajectories on a first day and a second day, with some aspect of the environment possibly changing from day one to day two. Each map presented the agent (initially on the left side of the map) with an obstacle in the middle (black square with white dots pattern in Fig. 13). The astronaut could take one of two paths (north or south corridors) to reach the right side of the map, where they could obtain one of the packages and reach the space station. The stimuli for the first day were always drawn from one of two possibilities: the agent could cross the bottom terrain and collect the bottom object, or cross the bottom terrain and collect the top object. We did not use paths where the agent crosses the top terrain because they are conceptually symmetrical to the two types of paths we used. The stimuli for the second day were drawn from one of eight options. These were generated by manipulating (i) the terrain that the agent took (top or bottom), (ii) the reward that the agent collected (top or bottom), and (iii) whether the location of the rewards was the same (same or inverted). This produced a total of 16 trials (see Appendix for details).

Procedure. The procedure was similar to Experiments 1–2. Participants first read a short tutorial that explained they would be watching astronauts navigate an alien planet on two different days and they were told their task would be to determine if the astronaut already knew the costs of traveling, and if she already knew how much she liked each care package. Participants were next presented with a simple questionnaire that ensured they understood the task. Only participants who responded all questions correctly were given access to the experiment (see Appendix for details). After the tutorial, participants completed the sixteen trials of the task. Participants responded three questions in each trial. The first question asked participants if the position of the care packages had switched on the second day (see Fig. 13a). This question ensured that participants noticed that the positions of the care packages had

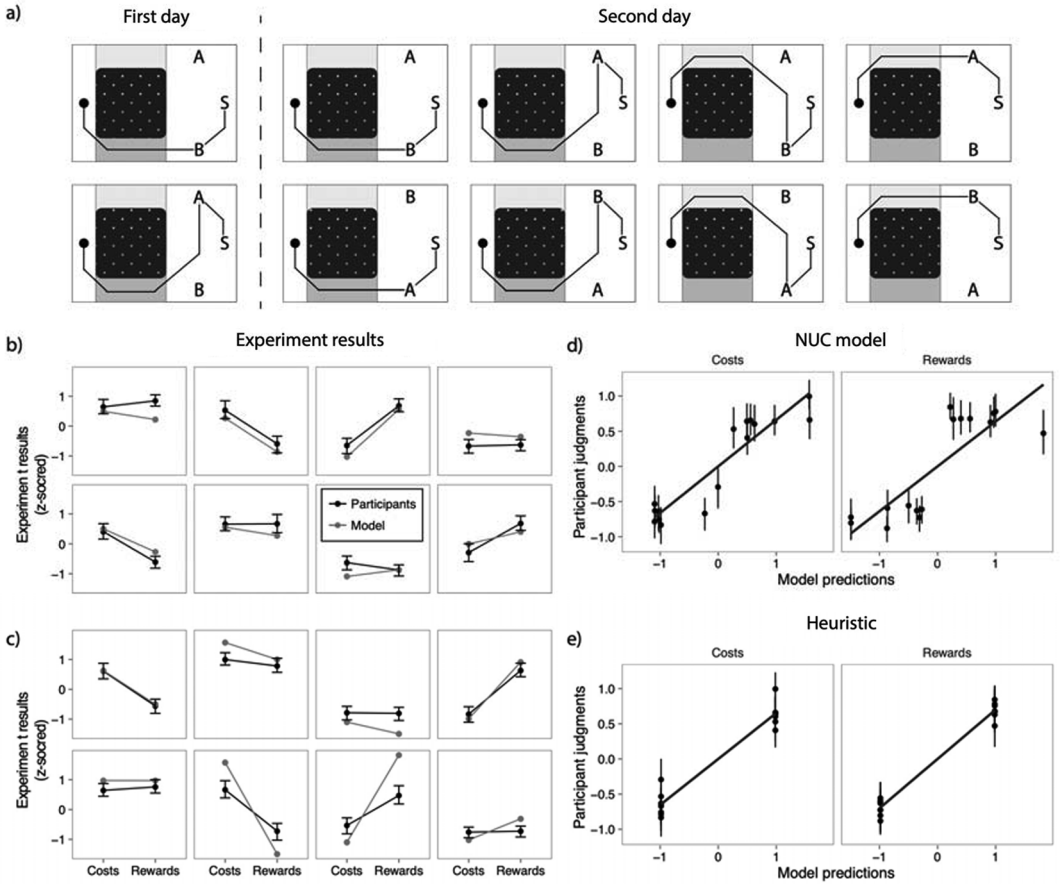

Fig. 13. Stimuli schematics and results from Experiment 5. Each trial consisted of two paths that the same astronaut took on different days. The first day consisted of one of two possible paths (left column in panel a), and the second day consisted of one of eight possible paths (right columns in panel a), producing a total of 16 stimuli. For the experiment, we generated up to twelve different version of each stimulus (by varying the colors of the terrain and colors of the care packages; see Appendix for different versions of each stimulus) and participants were shown a random one for each of the sixteen trials. Panels b-e show the results from Experiment 5. Each plot shows the model (gray) and participant's (black) likelihood that the agent was ignorant about the costs and rewards (z-scored). (b) Results from trials where the top image in panel a from Day 1 was used, (c) results from trials where the bottom image in panel a for Day 2 was used. (d) Main model results presented as scatterplots. Each dot represents a judgment where the x-axis shows the model predictions (z-scored) and the y-axis shows average participant judgments (z-scored). Vertical lines show 95% confidence intervals. (e) scatterplot equivalent to panel c bus using the heuristic predictions on the x-axis.

changed. Participants were not allowed to submit their responses unless that response was correct. The next two questions were "Did the astronaut already know the rewards on day 1?" and "Did the astronaut already know the costs on day 1?" Participants responded using a scale from 0 to 10 with labels "Definitely did not" on 0, "Maybe" on 5, and "Definitely did" on 10.

7.2. Results

Model predictions and participant judgments were z-scored in the same way as in Experiments 1 and 2 (see Results section of Experiment 1 for details). Fig. 13b-c show participant judgments along with model predictions with one plot per trial, ordered in the same way as the corresponding stimuli on Fig. 13a. Fig. 13d-e shows the same results as two scatterplots, one for costs and one for rewards, with model predictions on the x axis and participant judgments on the y axis. The Naïve Utility Calculus model showed a correlation of 0.93 (95% CI: 0.89–0.99) on cost knowledge inferences and a correlation of 0.83 (95% CI: 0.73–0.94) on reward knowledge inferences.

Next, we explored a simple heuristic where the agent was considered ignorant of the costs if she crossed different terrains on different days, and she was considered ignorant about the rewards if she chose different care packages on different days. This heuristic often makes the same predictions as our NUC model. However, the heuristic does not consider the possibility that an agent may have been initially ignorant about a certain cost or reward, but decided to not change their behavior in subsequent events either because they discovered the cost was low or because the reward was high. Fig. 13e shows the heuristic's predictions (x axis) plotted against participant judgments (y axis). The heuristic captured the broad structure of participant judgments with correlations of 0.97 (95% CI: 0.95–0.99) and 0.99 (95% CI:0.98–0.99) on costs and rewards with participant judgments, respectively. As Fig. 13e shows,

however, this heuristic predicts that participants should produce bimodal judgments with no graded information. Thus, to evaluate the heuristic further, we tested if each sub-region correlated with the Naïve Utility Calculus. On the cost knowledge inferences, we found a correlation of r = 0.63 (p = 0.09; n = 8 data points) on the low heuristic region and a correlation of r = 0.71 (p = 0.04; n = 8 data points) on the high heuristic region. On the reward knowledge inferences, we found a correlation of r = 0.51 (p = 0.20; n = 8 data points) on the low heuristic region, and a correlation of r = −0.66 (p = 0.08; n = 8 data points) on the high heuristic region. These results suggest that participant judgments were indeed influenced by attributions of how the agent's cost and rewards changed over time. If people relied on a heuristic like the one we described above, variability in responses should be experimental noise that does not correlate with the predictions of the Naïve Utility Calculus. At the same time, participant responses had substantially less variability than predicted by the NUC model, suggesting that when people make belief inferences from observable actions, they may be less able or less inclined to approximate ideal Bayesian inferences over the NUC model. We return to this point in the Discussion.

8. Experiment 6

Experiment 6 tests a final hypothesis of our proposal: if the Naïve Utility Calculus is instantiated as a generative model at the center of social reasoning, these inferences should also underlie how we reason about social goals. Past qualitative research with children has already shown that an expectation that agents maximize utilities underlies both social evaluations (Jara-Ettinger et al., 2015) and reasoning about communicative goals (Jara-Ettinger, Floyd, Huey, Tenenbaum, & Jara-Ettinger, 2019). Moreover, related computational work has shown that an expectation that agents move efficiently in space helps people infer whether agents intend to help or hinder (Ullman et. al., 2009). Experiment 6 adds to this body work by testing whether people's reasoning about social goals is sensitive to quantitative inferences about agents' unobservable costs. Experiment 6 was identical to Experiment 1a, with the difference that we replaced the care packages with stranded astronauts. Instead of asking participants to determine the astronauts' reward associated with the care packages, we asked participants to determine the astronauts' reward associated with helping the stranded astronauts.

8.1. Methods

Participants. 30 participants (mean age (SD): 34.46 years (6.55 years), range 23–48 years) from the US (as determined by their IP address) were recruited through Amazon's Mechanical Turk platform.

Stimuli. The stimuli were identical to the one from Experiment 1 with the exception that the "care packages" were replaced for "stranded astronauts" (see Fig. 3b).

Procedure. The procedure was identical to Experiment 1a with two exceptions. First, the tutorial explained that the protagonist could help stranded astronauts on her way to the space station, but we clarified that the stranded astronauts were not in mortal danger. Second, the reward question was changed to "how much does he like this stranded astronaut?" followed by a picture of the astronaut corresponding to the stimuli.

8.2. Results

Model predictions and participant judgments were z-scored in the same way as in Experiments 1–3 (see Results section of Experiment 1 for details). Participant judgments showed a correlation of 0.85 (95% CI: 0.77–0.94) and 0.93 (95% CI: 0.90–0.99) with our NUC model on the cost and reward dimensions, respectively. Fig. 14 shows the correlation between answers participants gave in Experiment 1 (collecting objects; x-axis) and in Experiment 6 (helping agents; y-axis). Participant responses correlated 0.94 (95% CI: 0.91–0.98) on costs and 0.98 (95% CI: 0.96–1.00) on rewards with responses obtained in Experiment 1, and these correlations were not reliably different between experiments (−0.045; 95% CI −0.11,0.02 and 0.01; 95% CI −0.03,0.03 for costs and rewards,

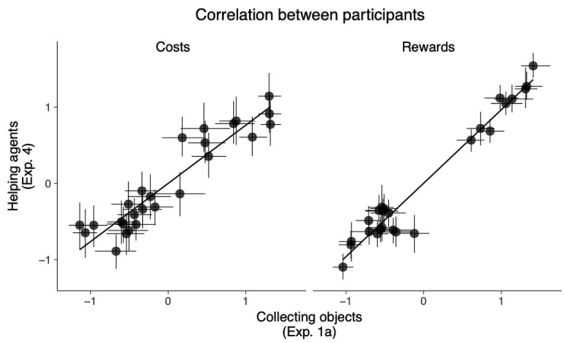

Fig. 14. Results from Experiment 6. The top row shows example of the stimuli. The bottom row shows participant judgments. Each dot represents a stimulus. The x-axis shows the judgments participants gave in Experiment 1 (where the protagonist was collecting objects) and the y-axis shows the judgments in the equivalent stimuli for the helping scenario.

respectively). This suggests that the same generative model of utility maximization explains how people infer agents' non-social rewards (such as collecting care packages) and social rewards (such as the motivation to help).

9. General discussion

We presented a computational model of a fundamental aspect of human social cognition: the ability to interpret other people's actions in terms of the motivating forces behind their goals. Our results converge to the idea that our fundamental ability to make sense of other people's actions is supported by a Naïve Utility Calculus—a mental model of others' choices and actions that works through the assumption that agents maximize utilities. Experiment 1 showed that people can infer an agent's unobservable costs and rewards given observable actions as predicted by the Naïve Utility Calculus, and that people are not only sensitive to qualitative differences (e.g. the choice of the terrain), but also to quantitative differences (e.g. the distance covered on each terrain), in a way that could not easily be explained by simpler heuristics or by a closely-related rate-based utility model. Experiments 2–3 provide further support to these conclusions, showing that the Naïve Utility Calculus also explains how participants infer the costs and rewards behind an agent acting in multiple events (Experiment 2), and it predicts participant confidence judgments with quantitative accuracy (Experiment 3). Experiments 4–6 suggest that the Naïve Utility Calculus is instantiated as a generative model that can predict a wide range of participants' inferences: predicting future goals in different situations (Experiment 4); judging agents' state of knowledge or ignorance from how they revise their behavior across situations (Experiment 5); and reasoning about an agent's socially-motivated goals (Experiment 6).

Across these studies, we also found some variability in model fits. Not every kind of judgment was equally well captured by our model. Still, the fact that a single model captures how participants make judgments in five different types of tasks across six experiments strongly suggests that a quantitative utility calculus is centrally involved in how we reason about others, and that this understanding is structured as a generative mental model that supports probabilistic inferences to reason flexibly about a wide variety of social situations.

9.1. Role of rationality and relation to other accounts

Our computational model invoked rationality in two different ways by performing rational inference over rational agents. Our use of rationality in observers was based on past computational work showing that people infer goals and preferences through Bayesian inference (Baker et al., 2009; Baker et al., 2017; Jern et al., 2017; Lucas et al., 2014), and our results provide further support to this idea.

Work in social psychology, however, has argued that people can be resistant to updating their impressions of others in light of new evidence (Petty, Tormala, Briñol, & Jarvis, 2006; Rydell & McConnell, 2006) and instead engage in motivated reasoning (Kunda, 1990). While it is possible that some representations of others (particularly implicit representations) are impervious to rational inference, some recent work suggests that people can indeed rapidly change their implicit representations when the evidence warrants it, and that some apparent failures to update beliefs are in fact consistent with rational inference under strong priors (see Ferguson, Mann, Cone, & Shen, 2019; Kim, Park, & Young, 2020 for reviews). Nonetheless, our work only shows that the building blocks of action understanding are quantitatively captured through Bayesian inference, but this does not imply that all social cognition works under the same inferential principles.

The second notion of rationality in our model—that agents act rationally to fulfill their goals—was historically contentious, with alternative accounts arguing that action understanding instead relies only on a capacity to map our own mental states and behaviors onto others (with rationality emerging only as a reflection of our own minds; Gordon, 1986; Thioux, Gazzola, Keysers, 2008; see Carruthers & Smith, 1996 for review). This family of accounts, known as simulation theory[2], however, has been limited in explanatory power and empirical support (Saxe, 2005; Hickok, 2009). To our knowledge, there are no current accounts that can explain empirical data on action understanding (Gergely & Csibra, 2003; Liu et al., 2017; Jern et al., 2017; Jara-Ettinger et al., 2016) without appealing to some notion of rational action.

A more intermediate stance comes from recent empirical evidence showing that people expect agents to develop habits which, in novel situations, can lead to inefficient behavior (Gershman, Gerstenberg, Baker, & Cushman, 2016; Goldwater et al., 2020). While these findings have been presented as evidence that we do not always expect agents to act rationally, the same results might support the opposite conclusion: the expectation for rational action may extend to cognitive costs, as habits allow agents to avoid having to repeatedly solve the same planning problem. Regardless of interpretation, however, an understanding of habit formation must emerge late in development: Infants expect agents to take novel efficient paths even after long periods of *habituation* to curved paths (Gergely et al., 1995; Gergely & Csibra, 2003), an effect that should not appear if they expected the agent to develop a habit. Nonetheless, the strongest support for the idea that we expect agents to act rationally, comes from our own and related work showing that computational models structured around this expectation explain a wealth of qualitative and quantitative data (Baker et al., 2009; 2017; Jara-Ettinger et al., 2016; Gergely & Csibra, 2003).

[2] Note that many computational models of action understanding, including our own, rely on some form of simulation, but they are not implementations of simulation theory. A central claim of simulation theory is that the simulation consists in considering how we ourselves would act in a given situation, rather than using simulations over a specialized generative model of rational action.

9.2. Model shortcomings, limitations, and possible extensions for future work

Overall, participants made many systematic graded judgments that the Naïve Utility Calculus model but not simpler heuristics were able to explain. An exception was Experiment 5, where we found conflicting evidence. Here participants had to infer whether an agent was knowledgeable about the costs and rewards, and our model captured people's judgments but predicted more variability than we observed. By contrast, a simple heuristic that mapped qualitative changes in behavior onto judgments about ignorance also captured people's judgments, but it predicted less variability than we observed. These results suggest that participants relied on a simpler computation that the one our model performs, but which is not as simple as our heuristic. Interestingly, from the Naïve Utility Calculus's standpoint, Experiment 5 is the most computationally-demanding task using a generative model. In contrast to the rest of our tasks, where our model had to infer only one cost and one reward function, in Experiment 5 our model had to infer a cost and reward function that could change over time based on where the agent navigated on the first day. Because of the increased complexity, and the salient behavioral features that reveal ignorance (such as changing one's behavior), people may rely on a simpler heuristic, a sampling-driven approach relying on only a few samples (e.g. Hamrick, Smith, Griffiths, & Vul, 2015), or some combination of both. More work is necessary to understand why people's judgments show traces of cost and reward inferences, but substantially less variability than we predicted.

Our model works by expecting agents to compute exact costs and rewards, and to consider every possible plan they may pursue. In this sense, our model is only a computational-level description of action-understanding that does not attempt to capture the algorithms or implementations that people use. At an algorithmic level, we expect people to rely on approximate cost and reward estimates rather than exact ones, especially in complex situations, and to not necessarily consider every possible available plan. Still the present algorithmic instantiation of our model may represent a step towards a more process-oriented account of people's judgments. A possible adaptation, for instance, would be to replace the exact planning algorithm we use with a sampling-based planner (similar to our sampling-based inference mechanism). This would produce approximate cost and reward estimates that are refined as more samples are drawn, yielding a spectrum of more or less accurate estimates that could vary across participants or task conditions. Alternatively, it is possible that mental state inferences rely on non-sampling approaches, such as those used in the inverse reinforcement learning literature (Collette, Pauli, Bossaerts & O'Doherty, 2017; Ziebart et al., 2008). More research is needed to explore these possibilities.

A further limitation of our model is that it assumes that the agent and the observer have perfect knowledge about the spatial layout and the location of the objects. In more realistic situations, agents can have incomplete knowledge about the world, and action-understanding must account for this. In related work, Baker et al. (2017) showed how partially-observable Markov Decision Processes (POMDPs), an extension of MDPs where agents can have uncertainty about the state of the world, can explain how people reason about other people's beliefs. In principle, it is possible to expand our model by replacing the goal-planning stage with POMDPs instead of MDPs so that planning and cost-estimates are sensitive to uncertainty in the world while leaving the main aspects of our framework untouched.

Another limitation of our work is that our experiments focused exclusively on understanding action in spatial contexts: agents moving around different terrains in pursuit of goals in different locations. Related work has shown that an expectation that agents maximize utilities also explains how we reason about agents making choices in non-spatial domains (Lucas et al., 2014; Jern et al., 2017). In particular, the Naïve Utility Calculus explains why, from infancy, we infer stronger preferences when agents choose rare items (Kushnir, Xu, & Wellman, 2010; Gweon, Tenenbaum, & Schulz, 2010; Wellman, Kushnir, Xu, & Brink, 2016; Hu, Lucas, Griffiths, & Xu, 2015; Fig. 15). The rarer an object is, the more likely that agents will have to incur additional costs in terms of distance, time, and attention to locate and retrieve it. As such, an agent that selected a rare object likely incurred a high cost, which, in turn, reveals a high reward. In related work (Jara-Ettinger, Sun, Schulz, & Tenenbaum, 2018), we have shown how a version of the model presented here can explain these judgments qualitatively, and presented a set of parametric studies testing the model's quantitative predictions for how these spatial and statistical factors tradeoff in judging agent's preferences (see Fig. 15d for a summary of some of these results).

Finally, here we focused on the role of costs and how they tradeoff with rewards. In more realistic scenarios, several additional complexities that we did not consider come into play. First, we expect agents to not only tradeoff rewards against costs, but also against risk. Risk can be seamlessly integrated into how our model estimates utilities, and in related work we have shown that this extension predicts people's inferences in situations with probabilistic outcomes with high accuracy (Jara-Ettinger & Gweon, 2017). The second complexity we did not consider is effort (Leonard, Lee, & Schulz, 2017; Leonard & Schulz, 2015). In our model, each combination of action and state is associated with a single cost and always produced a successful outcome. Intuitively, however, agents can choose how much effort to put into an action, and this affects the costs of taking that action and its probability of success. While our model can be extended so that agents can select how much effort to put into each action, we first need to characterize people's intuitions about the relation between effort, cost, and outcome. This is a direction we hope to explore in future work.

9.2.1. Naïve Utility Calculus in multi-agent settings

Our work focused on simple events where agents move around in space to collect objects and help static agents. How does the Naïve Utility Calculus support social reasoning in more complex situations where multiple agents are interacting with each other?

When reasoning about social interactions, the Naïve Utility Calculus can reveal when an agent is pursuing non-social, cooperative, and competitive goals, based on whether they are maximizing egocentric, group-level, or antagonistic utilities, respectively. Furthermore, social interactions often require combining our own costs and rewards with those of others (Kleiman-Weiner, Ho, Austerweil, Littman, & Tenenbaum, 2016; Török, Pomiechowska, Csibra, & Sebanz, 2019). To do this effectively, we must consider

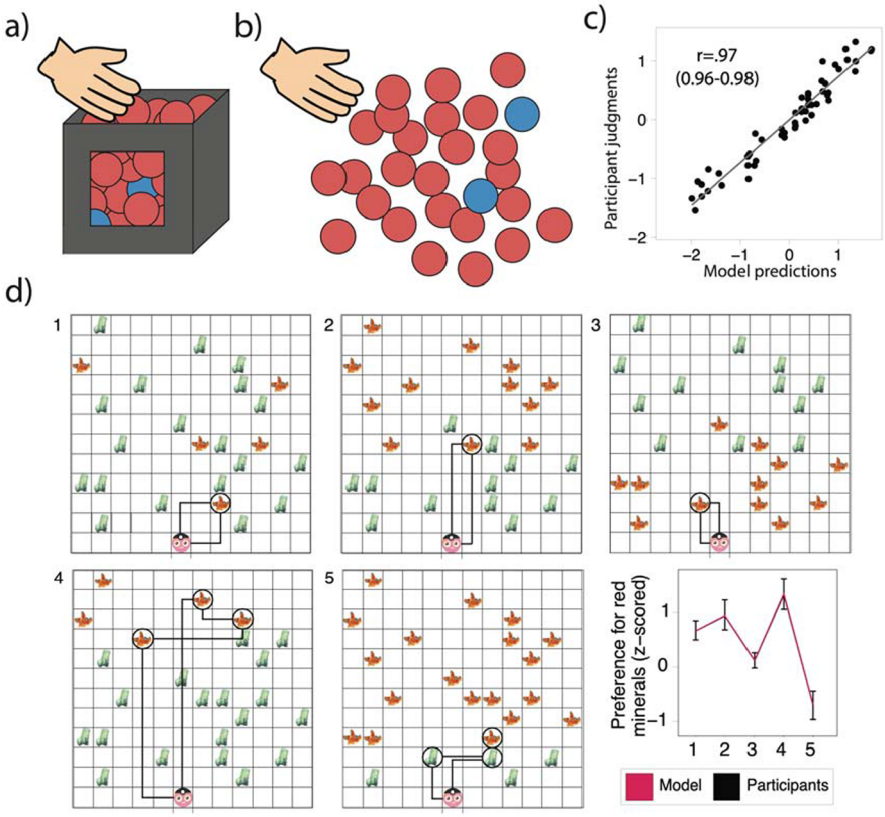

Fig. 15. (a) An example of a sensitivity to the sampling process scenario. If the agent were to draw a red (common; dark gray in grayscale) ball, we would infer a weak preference; if the agent were to draw a blue (rare; light gray in grayscale) ball, we would infer a strong preference. (b) Analogous situation to (a) with the contents spread out spatially. This shows that the rarity of a choice correlates with its cost: the rarer a choice is, the more likely that the agent has to incur additional costs to obtain it. (c) Utility maximization model predictions (x-axis) compared with participant judgments (y-axis) in a paradigm manipulating sampling information (from Jara-Ettinger, Sun, Schulz, & Tenenbaum, 2018). (d) Stimuli examples and results from the same experiment; in the plot, the x-axis shows the trial number and the y-axis shows the z-scored inferred relative preference for red (spikey) minerals over green (cylindrical) minerals. The black lines show participant judgments and vertical 95% confidence intervals and the red line shows the model predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

how other agents' costs differ from our own, and the Naïve Utility Calculus may also be instrumental in providing these estimates.

A further complexity is that social interactions often involve social costs and rewards, such as caring about social norms or other people's beliefs about us. In principle, our framework can be extended to account for any type of cost and reward in an event. However, how children learn the space of social costs and rewards that agents respond to is an open question. We will pursue these questions in future work.

10. Conclusion

People reason about others' goals and the causes behind them in terms of a Naïve Utility Calculus: We expect agents to maximize utilities, we conceptualize these utilities in terms of costs and rewards that vary across agents and situations, and these intuitions are instantiated as a generative model that supports a wide range of probabilistic inferences about agents' mental states, future actions, knowledge, and prosocial status.

In proposing a formal model of these capacities, we hope to have taken a step towards understanding the flexibility and explanatory depth of commonsense psychology. People's mental models of other agents lets them answer a seemingly unlimited number of questions (e.g. "Why did the agent choose what she did? Does she really like what she chose? Did she choose it because it was more convenient? Does she regret her choice? Was she pleasantly surprised? Did she underestimate her competence?"). Our work shows how decomposing agents' actions into costs and rewards can be the foundation enabling people to answer questions like these by mapping them onto inferences about the underlying cost and reward functions—inferences that can extend beyond the costs and rewards themselves to include judgments about competence, knowledge, future actions or social motivations. Our work also shows how a core set of representations and computations that let us reason about all agents—whether they are humans, other animals, or dots moving in a two-dimensional plane—allow us to build different expectations about different agents by understanding the ways

that agents vary in their costs and rewards, even as they always tend to act to maximize their subjective utilities.

At the same time, we must acknowledge that real-world behavior is far more complex than what we have attempted to model here, and the human capacity for action understanding can be correspondingly complex. People often act towards higher-order social goals (guided by friendship or reputational concerns), or act together to achieve shared or joint goals. People have incomplete knowledge about the world, about others, and even about their own costs and rewards. And we act in ways that cannot be reduced simply to moving through physical space. Nonetheless, our finding that human observers can perform rational quantitative cost and reward inferences in the simpler settings studied here suggests that these computations are fundamental, and at least are available for use in more complicated situations. The capacity to break down agents' behavior into fine-grained judgments about what they value and what they can do, with ease or with difficulty, may be central to our ability to thrive in a social world, enabling us to uncover what other people know, want and need, and guiding us in how to act in response.

Acknowledgments

We are grateful to Nancy Kanwisher, Rebecca Saxe, and Elizabeth Spelke for useful comments on the ideas behind this work. We thank two anonymous reviewers for critical feedback on the manuscript. This work was supported by the Simons Center for the Social Brain award number 6931582 and by a Google faculty research award. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF-STC award CCF-1231216.

Appendix A

Stimuli design, stimuli, and cover story

All Stimuli design details, experiment stimuli, and cover stories are available at: https://osf.io/uzs8r/.

Experiment 1

Correlation between Naïve Utility Calculus model and participant judgments in subsets of the data where the heuristic predicts that participants should give the same judgment for all trials. If the heuristic is correct, then any variability in participants should be noise and it should therefore not correlate with the Naïve Utility Calculus model. All twelve subsets of the data that contained at least six data points showed a positive correlation with the Naïve Utility Calculus:

Heuristic	Type	Experiment	r
-1.0329525	Cost	1a	0.9482711
-0.3840855	Cost	1a	0.9458871
-0.3355798	Cost	1b	0.6026243
0.9762086	Cost	1c	0.1827067
1.1745293	Cost	1b	0.4208993
1.1985170	Cost	1a	0.1851755
-1.1126973	Reward	1c	0.7385350
-1.0840273	Reward	1b	0.8902986
-0.7348469	Reward	1a	0.6474687
0.8345230	Reward	1c	0.8199009
0.8927284	Reward	1b	0.8879418
1.3063945	Reward	1a	0.9455087

References

Allais, M. (1953). L'extension des théories de l'équilibre économique général et du rendement social au cas du risque. *Econometrica, Journal of the Econometric Society,* 269–290.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*(3), 329–349.

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

Bettinger, R. L., & Grote, M. N. (2016). Marginal value theorem, patch choice, and human foraging response in varying environments. *Journal of Anthropological Archaeology, 42*, 79–87.

Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1655), 20130480.

Brown, R. (1986). Social Psychology, The Second Edition. Free Press.

Carruthers, P., & Smith, P. K. (Eds.). (1996). Theories of theories of mind. Cambridge University Press.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *Elife, 6*, Article e29718.
Constantino, S. M., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience, 15*(4), 837–853.
Csibra, G., Bıró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*(1), 111–133.
Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason'in infancy. Cognition, 72(3), 237-267.
Doan, T., Denison, S., Lucas, C. G., & Gopnik, A. (2015). Learning to reason about desires: An infant training study. In Proceedings of the Annual Meeting of the Cognitive Science Society.
Dvijotham, K., & Todorov, E. (2010, June). Inverse optimal control with linearly-solvable MDPs. In Proceedings of the 27th International Conference on International Conference on Machine Learning (pp. 335–342).
Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit first impressions can be updated. *Current Directions in Psychological Science, 28*(4), 331–336.
Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165–193.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences, 7*(7), 287–292.
Gershman, S. J., Gerstenberg, T., Baker, C. L., & Cushman, F. A. (2016). Plans, habits, and theory of mind. *PloS one, 11*(9).
Goldwater, M. B., Gershman, S. J., Moul, C., Ludowici, C., Burton, A., Killer, B., ... Ridgway, K. (2020). Children's understanding of habitual behaviour. *Developmental Science, e12951*.
Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2014). Concepts in a probabilistic language of thought. Center for Brains, Minds and Machines (CBMM).
Gordon, R. M. (1986). Folk psychology as simulation. *Mind & language, 1*(2), 158–171.
Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107*(20), 9066–9071.
Hamrick, J. B., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? The amount of mental simulation tracks uncertainty in the outcome. In Proceedings of the 37th Annual Conference of the Cognitive Science Society. Austin, TX.
Hawthorne-Madell & Goodman. (2015). So good it has to be true: Wishful thinking in theory of mind. Proceedings of the 37th Annual Conference of the Cognitive Science Society.
Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience, 21*(7), 1229–1243.
Hu, J., Lucas, C. G., Griffiths, T. L., & Xu, F. (2015). Preschoolers' understanding of graded preferences. *Cognitive Development, 36*, 93–102.
Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences, 29*, 105–110.
Jara-Ettinger*, J., Sun*, F., Schulz, L. E., & Tenenbaum, J. B., (in preparation). Sensitivity to the sampling process emerges from the principle of efficiency.
Jara-Ettinger, J., & Gweon, H. (2017). Minimal covariation data support future one-shot inferences about unobservable properties of novel agents. *Proceedings of the 39th annual conference of the Cognitive Science Society*.
Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20*(8), 589–604.
Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition, 140*, 14–23.
Jara-Ettinger*, J., Floyd*, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children believe that agents maximize expected utilities. Journal of Experimental Psychology: General.
Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B., & Schulz, L. E. (2019). Social pragmatics: Preschoolers rely on commonsense psychology to resolve referential underspecification. *Child Development*.
Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological science, 26*(5), 633–640.
Jern, A., & Kemp, C. (2011). *Capturing mental state reasoning with influence diagrams.* Cognitive Science Society.
Jern, A., & Kemp, C. (2014). *Reasoning about social choices and social relationships.* Cognitive Science Society.
Jern, A., Lucas, C. G., & Kemp, C. (2011). Evaluating the inverse decision-making approach to preference learning. In NIPS (pp. 2276–2284).
Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition, 168*, 46–64.
Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica. 47*(2), 263.
Kim, M., Park, B., & Young, L. (2020). The psychology of motivated versus rational impression updating. *Trends in Cognitive Sciences*.
Kleiman-Weiner, M., Saxe, R., & Tenenbaum, J. B. (2017). Learning a commonsense moral theory. *Cognition, 167*, 107–123.
Kleiman-Weiner, M., Ho, M. K., Austerweil, J. L., Littman, M. L., & Tenenbaum, J. B. (2016). January). Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. *CogSci.*.
Kryven, M., Ullman, T., Cowan, W., & Tenenbaum, J. B. (2015). Outcome of Strategy? A Bayesian Model of Intelligence Attribution. Proceedings of annual meeting of the cognitive science society.
Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin, 108*(3), 480.
Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*(8), 1134–1140.
Leonard, J., & Schulz, L. E. (2015). *If at First You Don't Succeed. The Role of Evidence in Preschoolers' and Infants' Persistence. Proceedings of the annual meeting of the cognitive science society*.
Leonard, J.A., Lee, Y., Schulz, L.E. (2017) Infants make more attempts to achieve a goal when they see adults persist. Science.
Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition, 160*, 35–42.
Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science, 358*(6366), 1038–1041.
Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., ... Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PloS one, 9*(3), Article e92160.
Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition, 120*(3), 403–411.
Mombaur, K., Truong, A., & Laumond, J. P. (2010). From human to humanoid locomotion—an inverse optimal control approach. *Autonomous Robots, 28*(3), 369–383.
Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology, 12*(7).
Ng, A. Y., & Russell, S. J. (2000). June). Algorithms for inverse reinforcement learning. *Icml, 1*, 663–670.
Pesowski, M. L., Denison, S., & Friedman, O. (2016). Young children infer preferences from a single action, but not if it is constrained. *Cognition, 155*, 168–175.
Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology, 90*(1), 21.
Ramírez, M., & Geffner, H. (2011, June). Goal recognition over POMDPs: Inferring the intention of a POMDP agent. In Twenty-Second International Joint Conference on Artificial Intelligence.
Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental Psychology, 33*(1), 12.
Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications, 9*(1), 1–12.
Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*, 995–1008.
Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences, 9*(4), 174–179.
Scott, R. M., & Baillargeon, R. (2013). Do infants really expect agents to act efficiently? A critical test of the rationality principle. *Psychological Science, 24*(4), 466–474.
Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17*(4), 443–464.

Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences, 110*(46), 18728–18733.
Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction(Vol. 1, No. 1). Cambridge: MIT press.
Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence, 112*(1–2), 181–211.
Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*(6022), 1279–1285.
Thioux, M., Gazzola, V., & Keysers, C. (2008). Action understanding: How, what and why. *Current Biology, 18*(10), R431–R434.
Török, G., Pomiechowska, B., Csibra, G., & Sebanz, N. (2019). Rationality in Joint Action: Maximizing Coefficiency in Coordination. *Psychological Science, 30*(6), 930–941.
Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In Advances in neural information processing systems (pp. 1874–1882).
Verma, D., & Rao, R. (2005). Graphical models for planning and imitation in uncertain environments. Technical Report 2005-02-01, Department of CSE, University of Washington.
Von Neumann, J., & Morgenstern, O. (1944). Theory of games and economic behavior.
Vos Savant, M. (1990). Ask Marilyn. *Parade Magazine, 15*, 17.
Wellman, H. M. (1990). The child's theory of mind.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 73*(3), 655–684.
Wellman, H. M. (2014). *Making minds: How theory of mind develops.* Oxford University Press.
Wellman, H. M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy*.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34.
Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). *How infants make sense of intentional action.* Intentions and intentionality: Foundations of social cognition149–169.
Ziebart, B. D., Maas, A. L., Bagnell, J. A., & Dey, A. K. (2008, July). Maximum entropy inverse reinforcement learning. In Aaai (Vol. 8, pp. 1433–1438).