# The naïve utility calculus: Joint inferences about the costs and rewards of actions

**Julian Jara-Ettinger (jjara@mit.edu), Laura E. Schulz (lschulz@mit.edu) & Joshua B. Tenenbaum (jbt@mit.edu)**
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, Cambridge, MA 02139 USA

**Abstract**

The understanding that agents have goals, and the ability to infer them, is fundamental in social cognition. However, much of our social understanding goes beyond goal attribution. Drawing on both behavioral studies throughout development, and on the limitations of past models, we propose that humans have a naïve utility calculus to reason about the costs and rewards underlying agents' goals. We show that the naïve utility calculus model, embedded in a Bayesian framework, can jointly infer the costs and rewards of agents navigating in complex scenarios. Using this model we test humans' ability to make quantitative cost-reward inferences in scenarios with various sources of costs and rewards. Our results suggest the naïve utility calculus model fits human inferences better than simple goal inference models.

**Keywords:** Bayesian modeling; Inverse planning; Naïve Utility Calculus; Social Cognition; Theory of Mind

## Introduction

Understanding that agents move to complete goals is at the heart of our social abilities and already at work in infancy (Woodward, Sommerville, & Guajardo, 2001). In addition to knowing that agents have goals, we also have expectations about *how* agents complete them. Developmental evidence suggests that humans expect agents to act efficiently (Scott & Baillargeon, 2013; Gergely & Csibra, 2003). This assumption, known as the *principle of efficiency*, enables humans to infer unobservable goals from observable behavior. The logic of this inference can be described and formalized using Bayesian inference, where the probability that an agent has goal $G$ given that they took actions $A$ is given by

$$p(G \mid A) \propto L(A \mid G) p(G). \tag{1}$$

Here, $L(A \mid G)$ is the likelihood that the agent would take actions $A$ if she had goal $G$, and $p(G)$ is the prior belief that the agent goal $G$. The principle of efficiency determines the likelihood function: The more efficiently the actions $A$ complete the goal $G$, the higher their likelihood (And therefore the higher the posterior probability that the agent has that goal).

This kind of inference, called inverse planning, was formally modeled by Baker, Saxe, & Tenenbaum (2009), using Markov Decision Processes (MDPs). In the MDP framework, the environment is modeled as a set of states, each with an associated utility (that can be positive or negative), which the agent can navigate by taking different actions (e.g., walk left, right, etc). With this formulation, it is possible to determine the sequence of actions that maximize an agent's utility as efficiently as possible. Using MDPs as a model for how agents act, goal inference can be formalized as inferring the unobservable utility function that is guiding the agent's actions. These models predict with high quantitative accuracy how adults infer goals in simple scenarios (Baker, et. al., 2009; 2011; Jara-Ettinger et al., 2012).

### Social reasoning beyond goal attribution

Despite the success of these models, the power of these inferences is limited.

**Explanatory limitations** To illustrate why, consider a simple example. A man is walking and reaches a fork on the road. The left path leads to a lake where he can swim, and the right path leads to his house. The man stops for a second and then takes the right path. The man's goal is immediately revealed after his first step, as he's taking an efficient path towards his house and an inefficient path towards the lake. However, this inference only tells us *what* the man is trying to achieve, but not *why*. The man may be going home because he doesn't like swimming, because he cannot swim, because he's too tired, or too hungry.

Models that infer the utility function will treat all the above explanations as being roughly equivalent, as they all reduce to a utility function with a higher value for being home than for going swimming. Intuitively, however, each statement tells us more about the man's psychological state and provides some insight into why swimming had a low utility. That is, rather than only reasoning about high utilities and associating them with goals, we are also sensitive to the costs and rewards underlying these utilities.

**Predictive limitations** Following on the past example, after the man arrives to his house, the predictive power of goal inference vanishes. We don't know what the man will do next, or even if he will have the same goal in the future. However, each explanation above boosts our predictive power. Knowing the costs and rewards underlying the man's goal allows us to reason about how his utilities may change over time. A tired man might choose to go swimming after taking a nap; an incompetent swimmer will not.

**Inferential limitations** Desires might be in direct conflict with each other (e.g., wanting to a cookie and wanting to lose weight), they might be too costly to obtain (e.g., buying a new car), or we may not know how to complete them (e.g., wanting world peace) (Moses, 2003). As such, agents have to compromise and tradeoff their true desires to choose a goal. Therefore, goals aren't always aligned with agents' true preferences. This makes it critical to distinguish between high utility states (what an agent wants to do at the moment) and high reward states (what an agent intrinsically likes). If your friend buys coffee next door you won't infer that she likes it better than the coffee sold across town, but if she goes all the way across town, you'll be confident she likes it better than the coffee from the local shop.

**Practical limitations** Standard goal attribution accounts assume that costs are identical for all agents. However, this is not the case. Consider this common scenario: Anne and Bob arrive to check-in at the airport and find that the entry is an empty zigzag pathway. Anne, who is six-years-old, takes *her* most efficient path towards the counter: ducking under the divisions. At the same time, Bob, who is 6' tall, takes *his* most efficient path towards the counter, by zigzagging through the path. If we assumed that both agents were acting efficiently with respect to the same objective costs we might infer that Bob is changing his goal at every bend in the zigzag path. Thus, to infer goals we need to understand that costs vary across agents, and we need to be able to infer them.

## The naïve utility calculus

In light of these limitations, our intuitive theory must also include some understanding of *how* costs and rewards jointly influence people's behavior. Recent developmental evidence suggest that we assume that agents estimate the costs and rewards associated with a goal, and chose what to do based on the difference of these two values: the utility.[1]

Preschoolers understand that costs and rewards vary across agents, and that these two determine the agent's utility, and thus their goals. Using this understanding, five- and six-year-olds can use knowledge about an agent's costs to infer their rewards, and, conversely, knowledge about an agent's rewards to infer their costs (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). At an even earlier age, two year-olds can estimate an agent's motivation to help using information about their costs (Jara-Ettinger, Tenenbaum, & Schulz,2015): When a competent and an incompetent agent refuse to help, toddlers infer that the competent agent was more likely to be unmotivated.

Intuitively, cost-reward tradeoffs happen in our everyday lives. We want to call our relatives but postpone it for weeks because we don't have time; we want to go to that nice restaurant downtown but end up going to the less desirable one near our house because it's closer, and we skip the best rides at theme parks because were not up for waiting in line. Formally, the utility for taking a sequence of actions $A$ to reach state $S$ is given by

$$U(S,A) = R(S) - C(A) \tag{2}$$

The higher a goal's utility, the more likely the agent will pursue it. Despite the simplicity, decomposing utilities into costs and rewards has powerful implications. Plans with high rewards and medium costs (e.g., doing something because you truly want it) are now different from plans with low rewards but even lower costs (e.g., doing something simply because it is convenient). Conversely, plans with low rewards and medium costs (e.g., foregoing something because you don't want it) are now different from plans with high rewards and even higher costs (e.g., foregoing something because it's too costly). However, the exact costs for different actions and the rewards for reaching different states vary across agents and are partially unobservable. Thus, for an observer to have the advantage of representing an agent's costs and rewards, they need to be able to infer them.

Despite the qualitative evidence for a naïve utility calculus early in development (Jara-Ettinger et al., 2014; 2015), the exact nature of these inferences, and the precision to which humans can make them, are open questions.

## Computational framework

To test people's ability to jointly infer an agent's costs and rewards, we implemented the naïve utility calculus model and a main alternative basic goal inference model (based on Baker, et. al., 2009). In addition, to get better insight into how each difference between the two models affects the cost-reward inferences, we implemented three additional intermediate models.

### Naïve Utility Calculus model sketch

This model is a direct extension of past goal-inference models (Baker, et al., 2009). However, rather than inferring the agent's utility function, we take the inference further and decompose the utility function into the underlying costs and rewards. This joint cost-reward inference can be seamlessly adapted into the inverse planning framework, where the probability that an agent who took actions $A$ has cost function $C$ and reward function $R$ is given by Bayes' rule:

$$p(C,R \mid A) \propto L(A \mid C,R)p(C,R)\,. \tag{3}$$

Here, the likelihood that the agent takes actions $A$ given their costs and rewards $C$ and $R$ is determined by the resulting utility function (Equation 2). That is, this model performs Bayesian inference over a generative planning model (formalized as a Markov Decision Process; See Baker, et al., 2009 for a detailed explanation of inverse planning through MDPs) by combining the cost and reward function to generate the utility function. Critically, the model understands that costs depend on the type of action (some actions are more costly than others) and on the agent (different agents incur different costs), and, similarly, that

[1] These types of models have been extensively studied as a theory for how humans produce behavior (Gilboa, 2010), but less

the rewards depend on the outcome (some outcomes are more rewarding than others) and on the agent (different agents place different rewards on the outcomes).

**Simple goal inference alternative model** As the main alternative we implemented a simple goal-inference model based on Baker, et al., (2009). Like the naïve utility calculus model, this model infers the unobservable utility function. However, rather than inferring an agent's costs, it assumes that all agents incur the same costs, independent of the action they take. Thus, this model is unable to infer agents' costs functions or to use them to infer the magnitude of the rewards.

## Intermediate accounts

**Competence inference model** This model extends the simple goal inference alternative model by allowing the costs to vary across agents. That is, this model assumes that agents incur a fixed cost for taking any action. However, it allows different agents to have different cost constants (their competence). As such, it understands that some agents may forego a high reward if the costs they would have to incur are too high. The difference between this model and the simple goal inference model quantifies the advantage an observer obtains by understanding that some agents are broadly more competent than others.

**Motivation inference model** This model is the complement of the competence inference model. As in the naïve utility calculus model, this model assumes that the cost for travelling depends on both the specific agent and the specific terrain. However, rather than inferring a separate reward value for each object, this model assumes that all objects have a constant reward value. Nevertheless, the model allows this value to vary across agents. Intuitively, the model attempts to explain agents' behavior by inferring their full cost function, and an overall level of motivation to complete goals. This model allows us to test if people's inferences can be explained by simply considering an agent's overall motivation to navigate the world and the cost they incur for navigating different types of terrains.

**Competence-motivation inference model** This last model assumes that agents' behavior is determined by two parameters: their overall competence and motivation. That is, the model assumes that each agent incurs a cost $c$ whenever it takes an action (regardless of the terrain) and obtains a reward $r$ whenever it collects an object (regardless of which object it collects). Although these two values are fixed for each agent, the model infers their specific value for different agents. This model, compared with the naïve utility calculus model enables us to quantify the inferential gain from giving the cost and reward functions more flexibility by allowing them to vary as a function of the objects and the terrains.

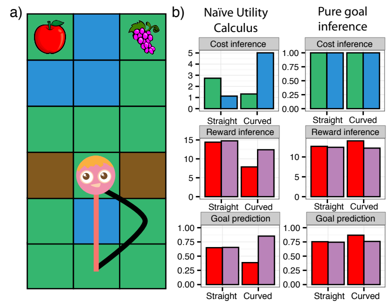


**Figure 1**. a) An agent moves from south to north towards two fruits. In the orange path, the agent moved in a straight line, while in the black path the agent circumvented the water. b) naïve utility calculus and simple goal inferences. Bars are color coded in accordance with the map.

To illustrate how the naïve utility calculus model and the simple goal-inference model differ, consider the sample path shown in Figure 1a. An agent is travelling from south to north, where he can pick up either, or both, of the fruits. The terrain consists of dense jungle (in green), water (in blue), and mountains (in brown). Figure 1b shows the two model's inferences for two potential paths. In the straight path (orange line) the agent travelled up north in a straight line, crossing the water. In the curved path (black line), the agent travelled up north circumventing the water. As the top row shows (Figure 1b), for the naïve utility calculus model, the straight path implies that the agent doesn't mind crossing water, and the curved path implies that he dislikes water. In contrast, the simple goal-inference model is unable to consider these differences. The second row shows each model's inferred reward functions. When the agent takes a straight path, both models infer that he probably likes both fruits. However, when the agent takes the curved path, the naïve utility calculus model now infers that the agent prefers grapes, while the simple goal inference model does not. This is consistent with the predictions about the agent's future actions (last row). Once again, the simple goal-inference model makes similar predictions for both paths. In contrast, the naïve utility calculus model infers that the agent is more likely to pick up the grapes when it observes the curved path, but not when it observes the straight path. Although simple, this example highlights how joint cost-reward inferences help overcome the limitations raised in the past section. The naïve utility calculus can infer *why* the agent circumvented the water, and it can use this knowledge to predict what the agent will do next. In contrast, the pure goal-inference model interprets all actions as attempts to reach the fruits through the shortest possible path.

# Experiment

To test people's ability to perform precise cost-reward inferences, we designed a simple experiment where participants were asked to infer the abilities and preferences of different agents navigating a grid world (as a static image) with three types of terrains and two types of objects.

## Design

The stimuli consisted of an 8x6 grid world with jungle, water, and mud (See Figure 2 for examples). Each stimulus contained the agent's starting point (which could be any of the four red squares shown in the examples in Figure 2), the end point (always located in the top left spot), two targets (located in any of the three possible locations shown in Figure 2; the apple and grape images were randomized across trials), and the agent's path. To generate the test stimulus we first ran 12,000 simulations (1,000 in each of the 12 possible worlds) of agents with random costs and rewards navigating the world (Cost and reward values were sampled from exponential distributions with parameters 0.1 and 10, respectively; these parameters were set qualitatively to ensure the simulations produced a wide range of paths). These simulations generated 189 unique paths. To reduce the stimuli size we first calculated each path's recoverability score, defined as the residual sum of squares (RSS) between the true parameters and the parameters inferred through Bayesian inference over the generative model (taking the posterior's expected value). Thus, paths with low recoverability indices had enough information for a rational observer to infer the underlying costs and rewards. Next, we calculated a discrepancy score for each alternative model, defined as the RSS between the naïve utility calculus predictions and the alternative model's predictions. Stimuli were reduced by removing all paths with a recoverability index greater than one, and then by selecting the 30 paths with the highest discrepancy score for each alternative model. The resulting 120 paths (30 for each of the four alternative models) reduced to 42 paths after removing duplicates. These 42 paths were thus ensured to contain enough information for observers to be able to make cost reward inferences (because they had a low recoverability index), and a high likelihood of helping us disambiguate between models (because they had a high discrepancy score). For each of the 42 paths we created an object version, where the map contained two fruits the protagonist could collect (See Figure 2), and a social version, where the map contained two agents the protagonist could help (The stimuli was otherwise identical). This allows us to test if humans make different cost-reward inferences when reasoning about social (helping someone) and non-social goals (collecting food). For instance, humans may infer a separate reward for each outcome in non-social goals (as the naïve utility calculus model does), but only an overall level of prosociality when reasoning about social goals (as the motivation inference model does).

## Participants

80 U.S. residents (as determined by their IP address) were recruited and tested through Amazon's Mechanical Turk platform (Mean age = 38.59 years. Min=19 years, max=68 years).

## Procedure

Participants were randomly assigned to the object (N=40 participants) or the social (N=40 participants) condition. In order to keep the experiment short, each participant only completed half (21) of the trials. These trials were selected by performing random splits, guaranteeing that each path was rated exactly 20 times in the social condition and 20 times in the object condition. Participants first completed a tutorial and a brief questionnaire to ensure they understood the task. Participants who responded one or more question incorrectly were automatically redirected to the beginning of the tutorial. Participants who responded all questions correctly were given access to the test stage. In each trial, participants saw a test path on the left side of the screen (See Figure 2 for examples; all images were static) and five slidebars on the right side of the screen. The first three slidebars asked about the agent's ability to navigate through each type of terrain (ranging from "Extremely exhausting" to "Extremely easy", with "average" in the middle) and the last two slidebars asked about the agent's strength of preference for each fruit, or about their motivation to help each stranded agent, depending on the condition (ranging from "Not at all" to "A lot" with no text in the middle).

## Results

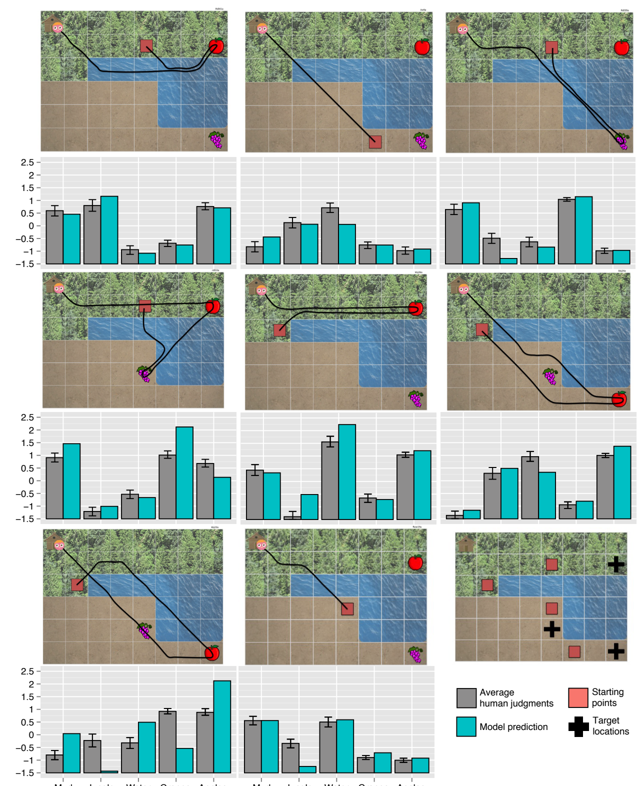


**Figure 2**. Example stimuli showing different starting points, object arrangements, and paths. Grey bars show average human judgments (z-scored per participant) with 95% confidence intervals. Teal bars show naïve utility calculus predictions.

As predicted, participants' average judgments were highly similar in the social and the object conditions (r=0.95; 95% CI: 0.93-0.97)[2], suggesting that people use the same type of reasoning when inferring an agent's social or non-social

[2] All reported confidence intervals were obtained through a basic non-parametric bootstrap.

rewards. In light of this, all further analyses were performed using the merged judgments from both conditions.

Figure 2 shows example paths with the naïve utility calculus inferences and the average human judgments. Although the model qualitatively matched human judgments, there were also high discrepancies. For example, in the path on the bottom left of Figure 2, humans inferred that the agent had a high reward for picking up both objects (or helping both agents). In contrast, the model inferred a high reward for the first target the agent reached and a substantially lower reward for the second object, as it was conveniently located on the agent's path towards the exit state (the top left of the map). This same path illustrates how the naïve utility calculus model showed more sensitivity to costs than humans did. At the beginning of the path the agent travelled north and moved two squares across the jungle before diving into the water. The model took this as strong evidence that the agent prefers navigating through the jungle relative to the other terrains, but humans did not.

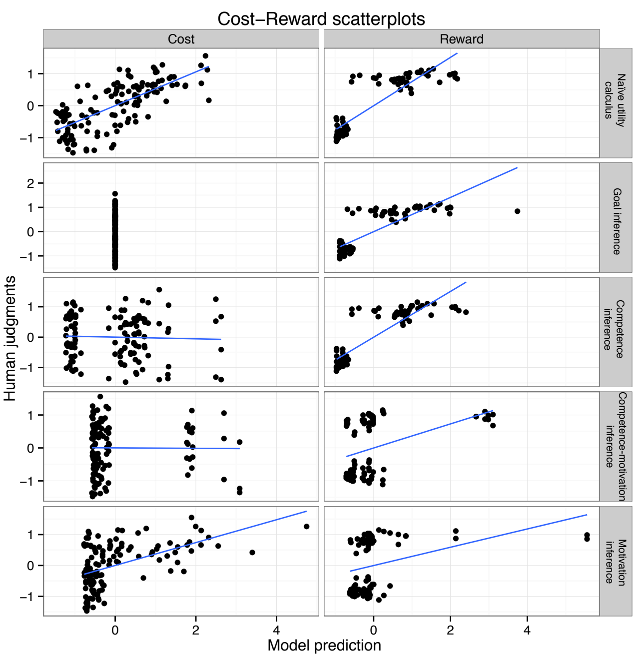


**Figure 3**. Scatterplot of model predictions (z-scored) compared to average human judgments. The x-axis shows the model predictions and y-axis shows the human (z-scored per participant) average judgments. The left column shows the cost inferences (three points per path) and the right column shows the reward inferences (two points per path). Each row shows a different model.

We next performed a quantitative model comparison by calculating each model's correlation with human cost and reward inferences (See Figure 3). To do this, each participant's data was standardized (z-scored) and then averaged. Similarly, each model's predictions were standardized (z-scored). On the cost dimension, the naïve utility calculus correlated the highest with human judgments (r=0.72; 95% CI: 0.65-0.79), followed by the motivation inference model (r=0.50; 95% CI: 0.40-0.61). The naïve utility calculus inferred the full reward function while the motivation inference model only inferred a single motivation parameter. Thus, this correlation difference (0.22; 95% CI: 0.09-0.34) suggests that inferring the reward function also helps recover the costs with more precision. The competence inference and the competence-motivation inference models both had correlations close to zero (r=-0.04 and -0.01, respectively. The 95% CI for both models was between -0.20 and 0.16), suggesting that humans do not treat costs as being uniform for each agent. Last, the simple goal inference alternative model makes no cost predictions and is thus incomparable on the cost dimension.

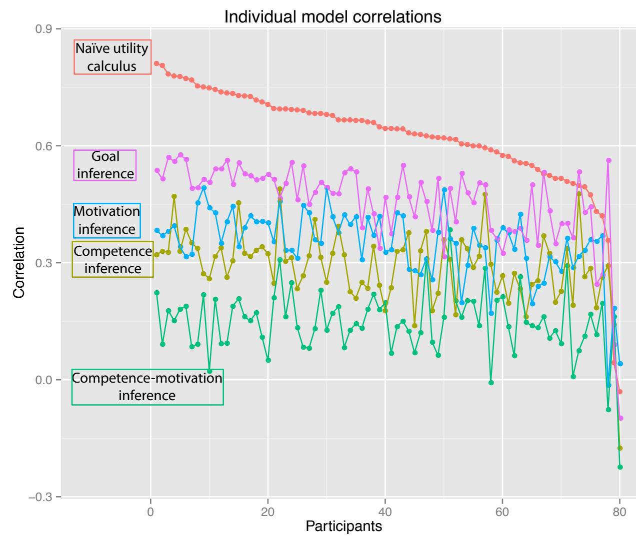


**Figure 4**. Individual model correlations with each participant. 93.75% of participants correlated best with the naïve utility calculus model. The x-axis shows all 80 participants. The y-axis shows each participant's correlation with each model. Participants are sorted by their correlation with the naïve utility calculus model. All model predictions were obtained prior to data collection and no individual parameters were fit.

On the reward dimension, the naïve utility calculus model showed the highest correlation (r=0.88; 05% CI: 0.83-0.93), but it was not reliably higher than the competence inference model (r=0.87; 95% CI: 0.82-0.93) or the simple goal inference model (r=0.82; 0.74-0.90) (95% CI difference between naïve utility calculus and competence inference and simple goal inference: -0.07-0.09 and -0.03-0.16, respectively). The motivation inference and motivation-competence inferences performed considerably worse (r=0.34 and 0.42, respectively; 95% CI: 0.44-0.68 and 0.32-0.57, respectively). Thus, our paradigm did not reveal any significant improvement in the ability to infer rewards by simultaneously inferring costs.

Last, we examined participants' individual performance by calculating their correlation with each model (See Figure 4). Because both cost and reward inferences were z-scored for participants and each model, we were able to calculate a joint cost-reward correlation score. All participants were correlated with the predictions generated from the model prior to data collection and no parameters were fit to individual participants. On average, participants had a correlation of 0.624 (95% CI: 0.60-0.66) with the naïve utility calculus model. Furthermore, 93.75% of participants

(N=75) showed the highest correlation with this model. Three out of the remaining five participants (6.25%) correlated better with the goal inference model and the other two participants correlated better with the motivation inference model (See Figure 4). This suggests that, although the naïve utility calculus model did not fit human inferences perfectly, it nevertheless clearly outperformed all other models at a global and individual level.

## Discussion

Here we proposed that the ability to reason about the costs and rewards underlying rational action is crucial for social reasoning. Inspired by developmental studies (Jara-Ettinger et al., 2015; Jara-Ettinger, Nate, Muentener, & Schulz, 2014) we implemented a formal model of the naïve utility calculus and tested its performance against human inferences.

Overall, the naïve utility calculus model outperformed the simple pure goal inference model as well as intermediate models both at a global level (averaging the responses of all participants) and at an individual level (correlating model predictions with individual participants). Importantly, the naïve utility calculus was able to infer the cost function in a quantitatively similar way to human's inferences (See Figure 3), which no other model was able to do. However, we also found unexpected results.

First, although the naïve utility calculus made better cost inferences compared to the other models, its reward inferences were matched by the simple goal-inference and the competence inference models. Thus, we failed to find evidence that the ability to infer an agent's costs helps to infer rewards with more precision. However, a closer look at the data (See Figure 3) suggests that, although the models showed a high numerical reward correlation, none of the models was able to predict human judgments with high accuracy. Critically, humans' reward inferences were bimodal, with participants mostly inferring that the agents' rewards took the highest possible value, or no value at all. In contrast, the naïve utility calculus model made graded predictions. One possibility is that humans were judging whether the agent placed a reward on the outcome or not, rather than inferring its exact magnitude. Further work is needed to determine if this effect is task specific or if it fundamentally reflects how humans make reward inferences.

In addition, our experiment only used complete paths. However, as Figure 1 shows, a significant advantage of jointly inferring the costs and rewards comes into play before the agent has completed their goal. Models that don't take into account an agent's costs assume the agent is always taking the shortest path towards their goal (which may not necessarily be the most efficient; see Figure 1) and thus can make incorrect inferences. As such, it is possible that the naïve utility calculus model would outperform the other models when making reward inferences in incomplete paths.

Importantly, participants performed identically in the object and the social conditions. This suggests that humans use the same kinds of inferences to reason about social goals. Having found overall support for human's naïve utility calculus, in future work we can bring this quantitative paradigm to study how humans make social and moral evaluations. Behavioral work suggests that the same kinds of inferences influence our social evaluations (Jara-Ettinger, Kim, Muentener, & Schulz, 2014; Jara-Ettinger, Tenenbaum, & Schulz, 2015). As such, models of people's quantitative cost-reward inferences may help us understand the precise computations underlying our social evaluations and moral judgments.

## Acknowledgments

We thank Samantha Floyd for useful comments and discussions. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF-STC award CCF-1231216, and by the Simons Center for the Social Brain.

## References

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

Baker, C. L., Saxe, R. R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the thirty-second annual conference of the cognitive science society* (pp. 2469-2474).

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

Gilboa, I. (2010). *Rational choice*. MIT Press.

Jara-Ettinger, J., Baker, C. L., & Tenenbaum, J. B. (2012). Learning what is where from social observations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 515-520).

Jara-Ettinger, J., Kim, N., Muentener, P., & Schulz, L. E. (2014) Running to do evil: Costs incurred by perpetrators affect moral judgment. In *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society*.

Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science.*

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*.

Moses, L. J. (2003). Some thoughts on ascribing complex intentional concepts to young children. *Intentions and intentionality: Foundations of social cognition,* 69-83.

Scott, R. M., & Baillargeon, R. (2013). Do infants really expect agents to act efficiently? A critical test of the rationality principle. *Psychological science*.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. *Intentions and intentionality: Foundations of social cognition*, 149-169.