Evaluating Critical Points in Trajectories

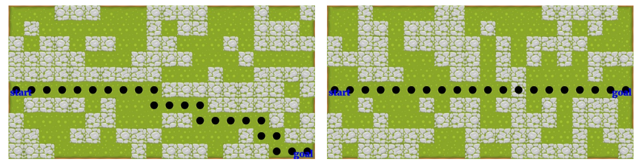
Shen Li*, Rosario Scalise*, Henny Admoni, Siddhartha S. Srinivasa, Stephanie Rosenthal

***Abstract*— People form beliefs about intentions and preferences of robots as they observe robot movement. However, robots rarely optimize their movement to allow people to easily determine state preferences. In this work, we define critical points along robot trajectories that convey information about state preferences: *inflection points* are changes in direction and *compromise points* are the relative proportion of preferred states to non-preferred ones. We contribute an approach for automatically generating trajectory demonstrations with specified critical points, and test observers' abilities to understand and generalize our robot's preferences based on our generated demonstrations. Our results show that inflection points helped participants understand state preference ordering and allowed them to more accurately predict paths through new environments, while compromise points hindered understanding. We conclude that robots should evaluate their trajectories for critical points to increase human observer understanding.**

I. Introduction

As robots perform tasks in human-occupied environments, people who observe them form beliefs about their behaviors [1]. Without insight into the robot's objective function or other information about how the robot behaves, people must derive their expectations from only the robot's motion within the context of the environment. These beliefs guide peoples' understandings and expectations of the robots as well as their interactions. If a person cannot understand why a robot planned its trajectory – even a successful one – they may not be able to predict its trajectory in a new environment.

Prior work has focused on using robot motion to effectively convey robot capabilities and goals [2, 3]. In contrast, we focus on using robot motion to convey its own objective function and show that it prefers to navigate through states with particular features. Consider the trajectory shown in Fig. 1a. It appears that the robot does its best to avoid rocks while navigating to the goal, implying it has a preference for traversing grassy states over rocky states. However, this trajectory could have also been generated by a robot with an objective function that has no preference for either terrain type if it arbitrarily chose where to turn. Similarly, a person observing the robot in Fig. 1b may be unclear about whether the robot has no terrain preference or a strong preference for grass. A person who does not understand the robot's objective function could be confused in a new environment when it does not plan a new trajectory that matches their expectations. We are interested in producing robot motion trajectories that help people understand the robot's feature preferences and that improve their ability to generalize that behavior to new environments.



(a) (b)

Fig. 1: Many possible objective functions could generate these trajectories.

Based on the observation that people assign rational meaning to agent actions [1, 4, 5] we define two types of critical points in a trajectory – *inflection points* and *compromise points* – as points that are information-rich and convey information about the relationship between the planned trajectory and the features in the environment. Fig. 1a is an extreme example of how inflection points (i.e., changes in direction) may lead an observer to infer preference for grass because the trajectory traverses only that terrain feature. The single rock compromise point in Fig. 1b may similarly lead an observer to believe there is no preference for grass over rocks when in fact all alternative paths have more rocks and therefore a lower overall value.

Our goal is to determine, in detail, the roles these kinds of points play in trajectories that lead to good understanding of robot behavior. Towards this, we conducted a large-scale study to systematically examine how varying the critical points in trajectories affects peoples' understandings of robot behaviors. We generated trajectories through synthetic environments according to different robot behaviors and showed them to people via Amazon Mechanical Turk (AMT). We conducted a within-subjects study in which we varied the parameterizations of the robot's reward function as well as the combinations of critical points along each trajectory and asked people to specify their understandings as well as generalize new plans in different environments. We show that people understand and can generalize the robot's terrain preferences more accurately as the number of inflection points increases and compromise points decreases within trajectories. However, when a robot has no preference for terrain types, the addition of either type of critical point within a trajectory reduces a participant's understanding.

We conclude that our critical points in trajectories do provide observers more information about a robot's state

*Both authors contributed equally.

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213
{shenli,robo,henny,siddh}@cmu.edu, srosenthal@sei.cmu.edu

 This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. DM17-0294

This work was (partially) funded by the DARPA SIMPLEX program through ARO contract number 67904LSDRP, National Institute of Health R01 (#R01EB019335), National Science Foundation CPS (#1544797), the Office of Naval Research, and the Richard K. Mellon Foundation.

preferences. A robot that can take these points into consideration while planning its trajectories can reduce observer uncertainty about its behavior while still acting optimally.

II. Related Work

The rationality principle states that people expect intentional agents, including robots, to choose actions to achieve their desires most efficiently [1, 4, 5]. However, humans often make errors in social judgment as they tend to attribute biased causes to others' behaviors (over-attribution effect) [6]. Towards the goal of helping people accurately understand robot behavior, prior research has focused on two areas: 1) autonomously recognizing planned behavior and 2) demonstrating plans to humans.

The goal of plan recognition is to passively infer agent goals and plans by partially observing agent actions [7]. Assuming all the possible plans are given by a plan library, some approaches are grammar-based parsing algorithms [8, 9] and Bayesian network inference algorithms [10, 11]. A more generative approach is to estimate agent goals without plan libraries through building agent decision-making models [7, 12–14]. Rather than observing actions to recognize plans, the goal of research in reward elicitation is to query human users for the utilities of actions that are performed to learn optimization functions [15]. A common approach is to ask users directly about their preferences to bound the estimations of their utility functions [15–17]. An indirect approach is to infer user reward function through observing user policy in response to incentives [18, 19]. In contrast to the automatic approaches in which a learner is trying to extract a human's plan or reward function, we are interested in changing the actions of a robot to allow a human to more easily learn the robot's reward function.

A growing area of research is focusing on ways to plan robot motion that is more interpretable or understandable to humans. Nikolaidis et al. have contributed action planning algorithms that allow their robot to reveal its capabilities adaptively through a game theoretic model of human expectations [2]. Other work has also developed expressive robotic lifting motions to help humans understand the weights of the objects that robots are manipulating [20, 21]. The ability of a person to recognize a robot's goals by observing its action execution also improves robot legibility [3, 22], predictability [23], acceptance [24], and naturalness [25], which are important for a human's recognition of robot tasks [26] and human-robot collaboration [27, 28]. However, the prior work aims to make *current* executed behavior and goals more understandable and does not focus on helping people more easily *predict* future actions and generalize current behavior to new environments.

Our approach to making robot behavior more understandable is to communicate the robot's preferences for different states or state features (its reward function) via its actions. Inspired by the idea that people attribute decision-making at critical points in behaviors to rationality [16], we propose critical points along a trajectory that could be more informative than others about the robot's preferences. We analyze how these critical points in a trajectory affect a person's understanding of the robot's reward function by systematically creating demonstration trajectories with particular sets of points. The demonstrations (either in simulations like ours or real robots like [29]) motivate people to observe new robot behaviors and infer the robot's preferences [18]. For each demonstration, we measured a person's ability to define the reward function and also generalize their observed trajectory to new environments.

III. Problem Formulation

We formulate our robots' behaviors as a standard Markov Decision Process which is a tuple of the form: $\{\mathcal{S}, \mathcal{A}, \mathcal{T}, R\}$.

This includes a set of world states $s \in \mathcal{S}$ with a single absorbing goal state $s_g \in \mathcal{S}$ and a set of robot actions $a \in \mathcal{A}$. The MDP has a deterministic state transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ and an immediate reward function $R : \mathcal{S} \to \mathbb{R}_+$. A robot behaves according to a deterministic policy $\pi : \mathcal{S} \to \mathcal{A}$. The optimal policy is denoted as π^* and describes the policy that maximizes the overall reward.

A trajectory $\xi(s_0|\pi) \in \Xi$ is defined as a sequence of states $[s_0, s_1, s_2, ..., s_g]$ where $\forall s_t \in \xi(s_0|\pi)$, $\mathcal{T}(s_{t-1}, \pi(s_{t-1})) = s_t$. The total reward of ξ is $R_\xi(\xi) = \sum_{s_t \in \xi} R(s_t)$. An optimal trajectory ξ^* is yielded by following π^*.

To ensure there are no cycles in a trajectory, there is one and only one $s \in \mathcal{S}$ such that $R(s) \geq 0$.

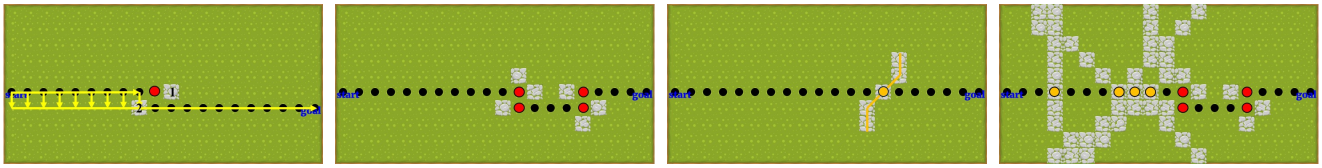
A. Experimental Setup

As an example domain, we consider a gridworld representation of a park which has a single terrain feature such as grass or rock assigned to each state (tile) on the grid.

- State $s \in \mathcal{S}$ is defined as $s = (x, y)$
- Action a is a 4-connected movement where $a \in \mathcal{A} = \{\to, \uparrow, \downarrow, \leftarrow\}$
- We define $\phi : \mathcal{S} \to \mathbb{N}_+^3$ as a mapping from states to features. $\boldsymbol{\phi}(s) = [\mathbb{1}_{\text{goal}}(s), \mathbb{1}_{\text{grass}}(s), \mathbb{1}_{\text{rock}}(s)] \in \{0,1\}^3$ subject to $\|\phi(s)\| = 1$, where each $\mathbb{1}(s)$ is an indicator function (e.g., $\mathbb{1}_{\text{grass}}(s) = 1$ if the tile type at s is grass and $\mathbb{1}_{\text{grass}}(s) = 0$ otherwise)
- We define $\mathcal{T}$ as a transition mapping with deterministic 4-connected movements within the gridworld.
- $\boldsymbol{\theta} \in \mathbb{R}^3$ are the weights for the feature vector $\boldsymbol{\phi}$. The reward for a state s with weights $\boldsymbol{\theta}$ is given by $R(\boldsymbol{\theta}, s) = \boldsymbol{\theta}^\intercal \boldsymbol{\phi}(s) \in \mathbb{R}$.
- When deriving the optimial policy, we break action ties with the ordering $[\to, \uparrow, \downarrow, \leftarrow]$.

IV. Critical Points of Trajectories

Depending on a robot's functional objective, the trajectory it follows can vary significantly. We characterize the information-rich states and actions within a trajectory as *critical points*. Based on the rationality principle, we focus on two types of critical points – *inflection points* in which people assign meaning to changing direction and *compromise points* in which a robot traverses over states with different features. Although this set of characteristics is not exhaustive, we believe it provides an effective starting



(a) (b) (c) (d)

Fig. 2: (a) generating 1 inflection point (red dot) by placing rock tiles at state 1 and 2 (b) 4 inflection points (red dots) (c) generating 1 compromise point (orange dot) by building a frontier (orange line segments) (d) 4 compromise points (orange dots)

point in analyzing trajectories. We will demonstrate that critical points can be beneficial in guiding the observer's understanding of robot behavior, or they can be detrimental to an observer's understanding, confounding their beliefs and leading to misinterpretation.

A. *Inflection Points*

Inflection points are defined as $s_t \in \xi(s_0|\pi)$ where the robot changes its direction. In other words, inflection points are all points at which the robot's action is not identical to its prior action (i.e., $\pi(s_{t-1}) \neq \pi(s_t)$). In Fig. 2a, an inflection point is indicated by the red dot where the robot moves up. This change of behavior gives people information about the robot's aversion towards the rock tile annotated as 1. In our park environment, inflection points come in pairs (e.g., the two inflection points in Fig. 2a) because the robot typically resumes moving rightward after changing direction.

B. *Compromise Points*

Compromise points are defined as states $s_t \in \xi^*(s_0|\pi^*)$ in which the myopic reward of entering s_t is not the maximum obtainable from s_{t-1}, yet the total reward for the trajectory is maximized. In particular, $\exists\ a_{t-1} \in \mathcal{A}, a_{t-1} \neq \pi^*(s_{t-1}), \mathcal{T}(s_{t-1}, a_{t-1}) = s'_t.\ s.t.\ R(s'_t) > R(s_t)$, but $R_\xi(\xi^*(s'_t|\pi^*)) < R_\xi(\xi^*(s_t|\pi^*))$.

The trajectory in Fig. 2c contains one compromise point (orange dot). To reach the goal, the robot must traverse a terrain feature ($e.g.$ rock) which incurs a higher cost than another possible terrain feature (grass) accessible from the previous state. Any attempt to move around the rock frontier would result in lower total trajectory reward compared to the straight path over the one compromise point.

V. Generating Demonstrations

We develop a method for synthesizing trajectories through environments that demonstrate the robot's reward function $R(\boldsymbol{\theta}, s)$ by changing $\boldsymbol{\phi}$ by iteratively inserting inflection and compromise points into the trajectory ξ^*.

A. *Inflection Points*

To create an inflection point at $s_i \in \xi^*(s_0|\pi^*)$, we can decrease the reward of s_{i+1} which alters $\pi^*(s_i)$ to avoid s_{i+1}. In Fig. 2a, grass is preferred and has lower cost than rock. To create an inflection point at s_i indicated as the red dot, we place a rock terrain tile at s_{i+1} annotated as state **1**.

One side effect of changing state **1** is that it might introduce multiple optimal policies yielding multiple optimal trajectories. The ambiguity of multiple optimal trajectories (or policies) can mislead people as it requires more complex reasoning to identify. One solution is to change some states to make all but one of the optimal trajectories sub-optimal. We treat this as a set cover problem. Universe U is the set of all the available optimal trajectories $U = \{\xi | \xi = \xi^* \leftarrow \pi^*\}$. $\forall s \in \xi \in U$, we define $subset(s) \subset U$ to include all the optimal trajectories that go through s (i.e., $subset(s) = \{\xi | s \in \xi \in U\}$). The family set contains all the $subset(s)$ (i.e., $set = \{subset(s) | s \in \xi \in U\}\ s.t.\ \bigcup_{ss \in set} ss = U$).

Our goal is to find the minimum number of states that all but one of the optimal paths include (i.e., to find the minimal set $cover$ subject to $\bigcup_{cc \in cover} cc = U \backslash \xi^{**}$ where $\xi^{**} \in U$ is the only path $s.t.\ \forall cc \in cover, \xi^{**} \notin cc$). $\forall subset(s) \in cover$, we can reduce $R(s)$ to make all the $\xi \in subset(s)$ sub-optimal and leave ξ^{**} the only optimal trajectory.

In Fig. 2a, there are 9 extra optimal trajectories available after changing states **1** (yellow arrows). By placing a rock terrain at state **2**, we could prevent the robot from moving downwards before reaching the red dot and make the trajectory indicated by black dots the only optimal trajectory. We generate 4 inflection points accordingly as shown in Fig. 2b.

B. *Compromise Points*

Similar to generating an inflection point, to generate a compromise point at $s_i \in \xi^*(s_0|\pi^*)$, we could decrease the reward of $s_{i+1}\ s.t.\ R(s_{i+1}) < R_{\max}$. But the difference is that now we want robots to keep $\pi^*(s_i)$ and head to s_{i+1} inevitably. Hence, we could decrease the rewards of a set of states in a neighboring area close to s_{i+1} to make it too costly for robots to detour around s_{i+1}. We could initiate the area as one state and iteratively increase its size until the new optimal trajectory passes through s_{i+1}. In each iteration, we could grow the area by making all the optimal trajectories which do not go through s_{i+1} become sub-optimal using the same technique we introduced in Sec. V-A.

In Fig. 2c, to create an compromise point at s_i (orange dot), we can build a frontier of states filled with rock terrain from top to bottom across the entire map (orange frontier). This frontier with low reward will force the robot to pass through s_{i+1} (the black dot on the right next to the orange dot). In our implementation, we use cubic Bezier curves [30] randomly generated through De Casteljau's Algorithm [31] to represent natural-looking frontiers. We generate 4 compromise points accordingly as shown in Fig. 2d.

C. *Extra Points*

We uniformly distribute different $\boldsymbol{\phi}$'s across our demonstration maps to so that all the maps are consistent with each other regarding the frequency of each feature. In our implementation, we ensure that each map contains 50% rocks and 50% grass adding complementary rock tiles to grass-dominant maps and vice versa. To make maps look natural, we place terrain types based on 2D Perlin noise [32–34]. Final maps and trajectories are shown in Fig. 3.

VI. Empirical Evaluation

We ran a study to test the effects of trajectories with different critical points on human understanding of robot terrain preferences. We presented participants with 16 different maps of "parks" with rock and grass terrain features, containing trajectories starting from the left and traversing to the right side of the park. We manipulated the number of critical points within trajectories as well as the actual terrain preference demonstrated in each map, and measured each participant's ability to predict the robot's preferences in a within-subject study design.

A. Independent Variables

We tested six terrain preference conditions and 10 no-preference conditions. The six preference conditions comprise all combinations of {0,2,4} inflection and {0,4} compromise points. The no-preference conditions are combinations of {0,2,4} inflection points, {0,4} compromise points, and {same, different} inflection point configurations[1].

Terrain Preferences. We compared trajectories through maps when there was a terrain feature preference versus when there was no preference between terrain features. We randomly selected half of the terrain preference conditions to prefer rock and half to prefer grass.

Inflection Points. Each demonstration trajectory had 0, 2, or 4 inflection points. Locations of the inflection points were randomly chosen along the path.

Compromise Points. We set the number of compromise points in each demonstration trajectory to be one of two values. When the reward function had preferences, these two values were {0, 4}. We were interested in observing the differences between having no compromise points versus having several points where the robot must '‘make a compromise" (which we chose to be 20% of the total trajectory length). When the reward function had no preferences, compromises could not technically occur. Therefore, we arbitrarily assigned a '‘simulated" preference and then divided the number of terrain features along the trajectory in the two levels: {50-50, 20-80}. The former level resulted in a trajectory where there was no preference illustrated by compromise points. The latter resulted in a trajectory where the robot simulated a compromise on 20% of the states.

Inflection Point Configuration. At each inflection point, there is a 'decision' corresponding to the change in direction. The robot's direction switches from continuing onto one tile (Fig. 4, B) to moving onto another tile (Fig. 4, C). We test whether human understanding changes if the terrain types of those tiles are the same (i.e., the robot chooses to turn from one grass tile to another grass tile) or different (i.e., the robot turns from a grass tile onto a rock tile). This condition is only tested when there is no preference in the terrain type.

[1]When there are no inflection points, there are no inflection point configurations, hence there are 10 'no preference' maps instead of 12.

B. Response Types

Sliders. We included a slider for each terrain feature type and labeled them {"Strongly Avoid", "Slightly Avoid", "Neutral", "Slightly Prefer", "Strongly Prefer"}. We asked participants to indicate the preference the robot had demonstrated for each terrain type using the sliders. Participants were free to place the sliders anywhere along the scale. We mapped their slider placements to a value between $[0, 1000]$, where 0 corresponds to "Strongly Avoid", 500 corresponds to "Neutral", and 1000 corresponds to "Strongly Prefer".

Text Free-Response. Participants were asked to explain the reasoning they believe the robot used as it planned its path through the map. Unlike the sliders, free response allows an unconstrained representation of the users' mental models of the robot behaviors. Due to space constraints, we do not present the results from the free response.

Drawing Trajectories. Last, we presented the participants with a new map (without a demonstration trajectory pre-drawn on it) and asked them to draw the trajectory they believed the robot would take if it were using the same reasoning to plan its new trajectory. Participants were required to start at a predefined point and could add 4-connected waypoints until reaching the goal position. Each map was generated to ensure it had a single optimal trajectory with respect to a fixed terrain preference. The maps were filled 50/50 with rock and grass tiles. In order to reduce the bias in our test maps, each participant received a randomized test map for each experimental condition. This measure allowed us to test participants' understanding of the robot's behaviors by comparing their drawn path to the optimal one.

Subjective Confidence. We asked participants to indicate on a 5-point Likert scale how confident they were that the trajectory they drew would be the one the robot would take.

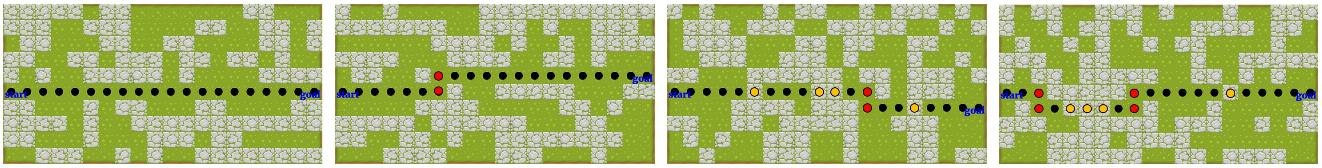
C. Study Deployment

We recruited 90 participants via Amazon Mechanical Turk. We used a within-subjects design where each subject was shown the total 16 conditions (6+10) in the same order. This order was pre-determined to ensure that no three consecutive conditions had the same terrain preference, which allowed us to avoid users inferring incorrectly based on coincidental patterns. Upon completion of the study, we collected demographic information from participants, including their age, gender, occupation, primary language, and experience with robots, video games, and RC-cars. We also asked for general comments as well as how difficult they found the tasks. Due to space constraints, these results are omitted.

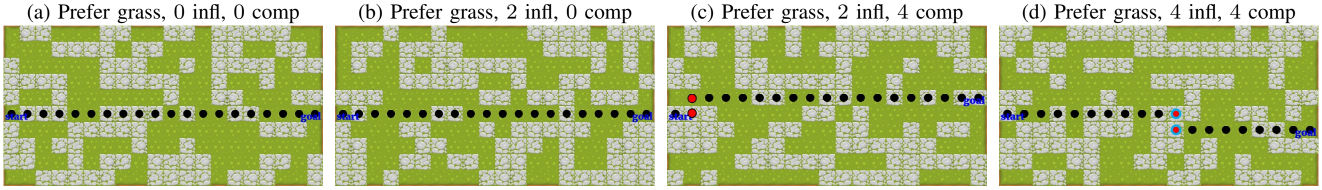
VII. Results

A. Dependent Variables

Our measures of accuracy in understanding robot preferences are based on the drawn trajectories, sliders, and subjective ratings of confidence.



(a) Prefer grass, 0 infl, 0 comp (b) Prefer grass, 2 infl, 0 comp (c) Prefer grass, 2 infl, 4 comp (d) Prefer grass, 4 infl, 4 comp



(e) No pref., 0 infl, 0 comp (f) No pref., 0 infl, 4 comp (g) No pref., 2 infl, different, 0 comp (h) No pref., 2 infl, same, 0 comp

Fig. 3: Robot preference type, number of inflection points (red dots), inflection point configuration ("different" = red dots with black circles, "same" = red dots with blue circles), number of compromise points (orange dots) for demonstration examples

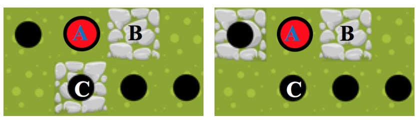


Fig. 4: (left) an inflection point with 'same' configuration (right) an inflection point with 'different' configuration

The *optimality ratio* $= |\frac{\text{total cost of optimal trajectory}}{\text{total cost of drawn trajectory}}| \in [0, 1]$. As people understand the robot reward function more accurately, the optimality ratio increases.

We assume that people use the distance between the grass and rock slider placements to indicate their certainty about inferring the robot preferences. We map the distance between the grass and rock slider placements to *preference range* $\in [0, 2000]$. A value of 0 corresponds to the user inferring no preference between the grass and rock terrains while a value of 2000 corresponds to the user inferring a difference with a high certainty, regardless of what the robot actually prefers.

We use *subjective confidence* $\in \{1, 2, 3, 4, 5\}$ to represent the user's self-reported confidence in understanding robot reasoning, with higher values indicating more confidence.

B. Hypotheses

H1 Preference demonstrations: increasing the number of inflection points will increase optimality ratio, preference range, and subjective confidence.

H2 Preference demonstrations: increasing the number of compromise points will decrease optimality ratio, preference range, and subjective confidence.

H3 No preference demonstration: increasing the number of inflection points will decrease optimality ratio, preference range, and subjective confidence.

H4 No preference demonstration: increasing the number of compromise points will decrease optimality ratio, preference range, and subjective confidence.

H5 No preference demonstration: the optimality ratio, preference range, and subjective confidence are lower when each inflection point has a different configuration than when each inflection point has the same configuration.

C. Results

1) Preference: **Optimality Ratio.** We use a two-way repeated measures ANOVA to find the effect of inflection points and compromise points on optimality ratio (Table I).

The number of inflection points has a significant effect on the optimality ratio $(F(2, 178) = 46.159, p < 0.001)$. Post hoc analysis with a Bonferroni adjustment reveals that the optimality ratio is significantly increased from 0 to 2 $(p < 0.001)$ and from 0 to 4 $(p < 0.001)$, but not from 2 to 4 inflection points $(p = 0.052)$, though it is close (Fig. 5a). This suggests that inflection points help users understand robot preferences. For example, it is easier for people to understand that the robot prefers grass over rock terrains by looking at Fig. 3b than Fig. 3a. Additionally, in these maps there is little benefit to showing more than 2 inflection points. For example, it is more difficult for people to understand that the robot prefers grass over rock terrains by looking at Fig. 3b than Fig. 3d although Fig. 3d has more inflection points. The first part of H1 is supported.

Increasing compromise points from 0 to 4 significantly decreases the optimality ratio $(F(1, 89) = 74.476, p < 0.001)$ (Fig. 5b). This suggests that path entropy hinders people's understanding of robot preferences. For example, it is easier for people to understand that the robot prefers grass over rock terrains by looking at Fig. 3b than Fig. 3c. The first part of H2 is supported.

There is a significant interaction between the numbers of inflection and compromise points on optimality ratio $(F(2, 178) = 5.291, p = 0.006)$. When there are no compromise points, there is no significant difference between 2 and 4 inflection points $(p = 0.730)$. However, when there are 4 compromise points, optimality ratio is significantly increased from 2 to 4 inflection points $(p = 0.001)$ (Fig. 5c). This indicates that as the number of compromise points increases, people need more inflection points to mitigate their confusion about the compromise points. For example, it is easier for people to understand that the robot prefers grass over rock terrains by looking at Fig. 3d than Fig. 3c.

Preference Range. We used a two-way repeated measures ANOVA to determine the effects of inflection points and compromise points on preference range (Table I). The num-

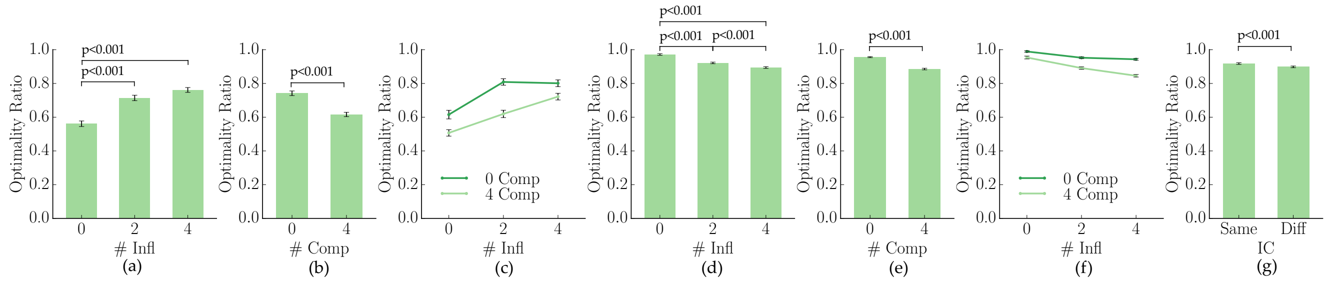


Fig. 5: When there is a preference, optimality ratio vs (a) the number of inflection points (b) the number of compromise points (c) the interaction between the number of inflection points and compromise points. When there is no preference, optimality ratio vs (d) the number of inflection points (e) the number of compromise points (f) the interaction between the number of inflection points and compromise points (g) inflection point configuration

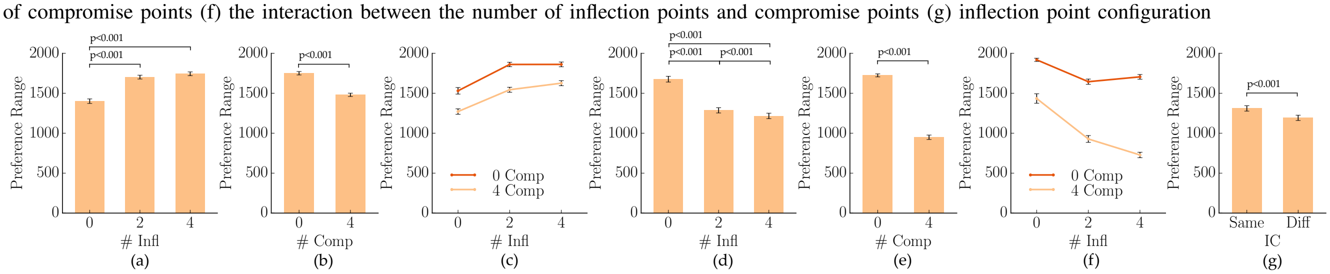


Fig. 6: When there is a preference, preference range vs (a) the number of inflection points (b) the number of compromise points (c) the interaction between the number of inflection points and compromise points. When there is no preference, preference range vs (d) the number of inflection points (e) the number of compromise points (f) the interaction between the number of inflection points and compromise points (g) inflection point configuration

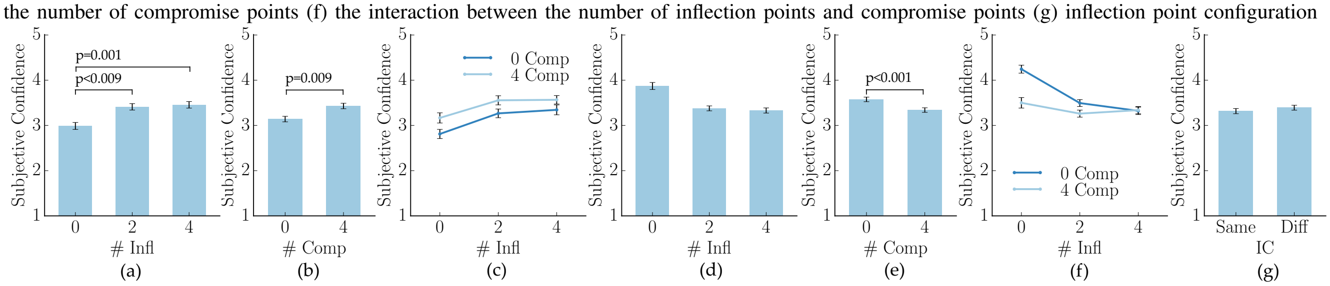


Fig. 7: When there is a preference, subjective confidence vs (a) the number of inflection points (b) the number of compromise points (c) the interaction between the number of inflection points and compromise points. When there is no preference, subjective confidence vs (d) the number of inflection points (e) the number of compromise points (f) the interaction between the number of inflection points and compromise points (g) inflection point configuration

ber of inflection points has a significant effect on preference range ($F(2, 178) = 65.759, p < 0.001$). A post hoc analysis with a Bonferroni adjustment reveals that the preference range is significantly increased from 0 to 2 ($p < 0.001$) and from 0 to 4 ($p < 0.001$), but not from 2 to 4 ($p = 0.385$) inflection points (Fig. 6a). This suggests that more inflection points lead to greater certainty about the robot's preference. Similar to optimality ratio, increasing beyond 2 inflection points does not improve preference range. The second part of H1 is supported. Preference range is also significantly decreased from 0 to 4 compromise points ($F(1, 89) = 91.050, p < 0.001$) (Fig. 6b). The second part of H2 is supported. There are no other significant effects on preference range.

Subjective Confidence. To measure the effect of inflection and compromise points on the Likert scale responses for subjective confidence, we ran a generalized ordinal logistic model and estimated the model parameters through a generalized estimating equation (GEE) with AR(1) covariance structure (Table I). Subjective confidence significantly increased from 0 to 2 ($p = 0.009$) and from 0 to 4 ($p = 0.001$), but not from 2 to 4 ($p = 0.907$) inflection points (Fig. 7a). This suggests that inflection points help people feel more confident about their evaluations, but that increasing beyond 2 inflection points does not necessarily lead to more confidence. The third part of H1 is supported. Subjective confidence is significantly increased from 0 to 4 compromise points ($p = 0.009$) (Fig. 7b). This suggests that path entropy decreases users' feelings of confidence in their evaluations. Interestingly, the third part of H2 is rejected. There are no other significant effects for subjective confidence.

2) No Preference: Analysis for no preference maps follows the analysis for preference maps above. Results for inflection point configuration are only available for demonstrations with 2 or 4 inflection points, since 0 inflection points mean there cannot be inflection point configurations.

Optimality Ratio. We conducted a three-way repeated measures ANOVA to determine the effect of inflection points, compromise points, and inflection point configuration on optimality ratio (Table II).

Number of inflection points significant affects optimality ratio ($F(2, 178) = 42.050, p < 0.001$). Post hoc analysis with a Bonferroni adjustment reveals that optimality ratio is significantly decreased from 0 to 2 ($p < 0.001$), from 0 to 4 ($p < 0.001$), and from 2 to 4 ($p < 0.001$) inflection points (Fig. 5d). This suggests that people's ability to identify the robot's true preferences continues to decrease as inflection points are added. For example, it is easier for people to understand that the robot has no preference by looking at Fig. 3e than Fig. 3g. The first part of H3 is supported.

Optimality ratio is significantly decreased from 0 to 4 compromise points $(F(1, 89) = 62.649, p < 0.001)$ (Fig. 5e), indicating that path entropy reduces people's ability to identify the robot's true preference. For example, it is easier for people to understand that the robot has no preference from Fig. 3e than Fig. 3f. The first part of H4 is supported.

There is a significant interaction between the numbers of inflection and compromise points on optimality ratio, $F(2, 178) = 12.652, p < 0.001$. When the number of compromise points is high, the optimality ratio is significantly decreased from 2 to 4 inflection points ($p < 0.001$), while when number of compromise points is low, there is no significant difference ($p = 0.883$) (Fig. 5f). This indicates that when there are many compromise points, more inflection points exacerbates the detrimental effect of compromise points on optimality ratio, while when the number of compromise points is low, the detrimental effect is gone.

Optimality ratio is significantly higher when inflection points have the "same" configuration than when they have a "different" configuration $(F(1, 89) = 12.793, p = 0.001)$ (Fig. 5g). This indicates that for maps without a preference, inflection points that move to the same type of terrain better reveal the robot's true (lack of) preference. For example, it is easier for people to understand that the robot has no preference by looking at Fig. 3h than Fig. 3g. The first part of H5 is supported. No other significant results were found.

Preference Range. We use a three-way repeated measures ANOVA to determine the effect of the number of inflection points, compromise points, and inflection point configuration on preference range (Table III). The number of inflection points has a significant effect on preference range $(F(2, 178) = 67.728, p < 0.001)$. Post hoc analysis with a Bonferroni adjustment reveals that preference range is significantly decreased from 0 to 2 ($p < 0.001$) and from 0 to 4 ($p < 0.001$), but not from 2 to 4 inflection points ($p = 0.069$) (Fig. 6d). The second part of H3 is supported. Preference range is also significantly decreased from 0 to 4 compromise points $(F(1, 89) = 181.118, p < 0.001)$ (Fig. 6e). The second part of H4 is supported.

There is a significant interaction between the numbers of inflection points and compromise points on preference range $(F(2, 178) = 18.848, p < 0.001)$. When there are 4 compromise points, preference range is significantly decreased from 2 to 4 inflection points ($p = 0.003$), while when there are 0 compromise points, there is no significant difference ($p = 0.611$) (Fig. 6f). This indicates that inflection points have a detrimental effect on preference range only when they are exacerbated by compromise points, but that without the compromise points there is no detrimental effect.

Preference range was significantly decreased from "same" to "different" inflection point configuration $(F(1, 89) = 13.802, p < 0.001)$ (Fig. 6g). This indicates that for maps without a preference, the preference range is lower when all inflection points have the "different" configuration than when the same number of inflection points have the "same" configuration. The second part of H5 is supported. No other significant differences are found.

Subjective Confidence. To determine the effect of inflection points, compromise points, and inflection point configurations on subjective confidence, we conducted a generalized ordinal logistic model and estimated the model parameters through a generalized estimating equation (GEE) with AR(1) covariance structure (Table IV). There is no significant effect of inflection points on subjective confidence (Fig. 7d). People are not significantly less confident about inferring the robot reasoning when dealing with demonstrations with more inflection points. The third part of H3 is rejected. Subjective confidence is significantly decreased from 0 to 4 compromise points ($p < 0.001$) (Fig. 7e). People are less confident about the robot's reasoning when dealing with demonstrations with more compromise points. The third part of H4 is supported. There were no significant effects of inflection point configuration on subjective confidence (Fig. 7g). The third part of H5 is rejected.

VIII. DISCUSSION

People derive expectations about robot behavior by observing robot trajectories. Our work serves as a basis for enabling robots to use the trajectories they take to convey information about their reward functions. In this work, we introduce the concept of critical points and give two examples – inflection points and compromise points. Using these, we develop a method for systematically generating trajectories that possess the critical points we specify. We then test how trajectories with varying combinations of critical points affect human understanding of robot reward functions. We show that inflection points can have different effects on human understanding depending on whether a robot's reward function has particular terrain feature preferences or not. Specifically, when there is a preference for terrain features, adding inflection points improves human understanding, while when there is no preference, adding inflection points hinders understanding. In both cases, increasing the number of compromise points decreases human understanding of the robot's preferences.

Interestingly, our results showed that the subjective confidence did not increase with fewer compromise points as we expected. Future work is needed to understand why this is the case. For example, it is possible that if participants never saw the robot navigate over a rock, they would not be confident about what would happen if it *had* to navigate over a rock.

Additionally, our results showed that there was a significant effect of one pair of inflection points but no benefit to the second pair of inflection points suggesting that there is a "law of diminishing returns" in information conveyed by inflections. Because we only investigated two terrain types, one pair of inflection points is all that is necessary to indicate which terrain type is preferred. More work is needed to investigate whether our finding holds for more complex environments. For example, while we believe that one inflection point is needed to show relative preference between pairs of features, it is unclear whether the complexity of the path will overwhelm an observer rather than help them.

Finally, our study was performed in an online study and not on a real robot. We acknowledge that it may be difficult to modify real environments in order for optimal trajectories to include critical points. In environments where a real robot cannot demonstrate its reward function by adding inflection points, for example, it may be possible for the robot to display a simulated environment with a trajectory (such as those we generated) to efficiently teach an observer about its preferences. Another option may be to demonstrate a non-optimal path that has more critical points. Future work is needed to understand whether our findings translate to real robots in real environments, and also whether other methods of demonstration are effective.

References

[1] G. Gergely, Z. Nádasdy, G. Csibra, and S. Bíró, "Taking the intentional stance at 12 months of age," *Cognition*, vol. 56, no. 2, pp. 165–193, 1995.

[2] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *Proceedings of HRI*, 2017.

[3] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *Proceedings of HRI*. ACM/IEEE, 2013, pp. 301–308.

[4] D. C. Dennett, *The intentional stance*. MIT press, 1989.

[5] K. Kamewari, M. Kato, T. Kanda, H. Ishiguro, and K. Hiraki, "Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion," *Cognitive Development*, vol. 20, no. 2, pp. 303–320, 2005.

[6] P. E. Tetlock, "Accountability: A social check on the fundamental attribution error," *Social Psychology Quarterly*, pp. 227–236, 1985.

[7] M. Ramırez and H. Geffner, "Plan recognition as planning," in *Proceedings of IJCAI*. Morgan Kaufmann Publishers Inc., 2009, pp. 1778–1783.

[8] D. V. Pynadath and M. P. Wellman, "Generalized queries on probabilistic context-free grammars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 65–77, 1998.

[9] C. W. Geib and R. P. Goldman, "A probabilistic plan recognition algorithm based on plan tree grammars," *Artificial Intelligence*, vol. 173, no. 11, pp. 1101–1132, 2009.

[10] H. H. Bui, "A general model for online probabilistic plan recognition," in *Proceedings of IJCAI*, vol. 3, 2003, pp. 1309–1315.

[11] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Artificial Intelligence*, vol. 171, no. 5-6, pp. 311–331, 2007.

[12] M. Ramírez and H. Geffner, "Probabilistic plan recognition using off-the-shelf classical planners," in *Proceedings of AAAI*. AAAI Press, 2010, pp. 1121–1126.

[13] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

[14] M. Ramírez and H. Geffner, "Goal recognition over pomdps: Inferring the intention of a pomdp agent," in *Proceedings of IJCAI*. AAAI Press, 2011, pp. 2009–2014.

[15] K. Regan and C. Boutilier, "Regret-based reward elicitation for markov decision processes," in *Proceedings of Uncertainty in Artificial Intelligence*. AUAI Press, 2009, pp. 444–451.

[16] U. Chajewska, D. Koller, and R. Parr, "Making rational decisions using adaptive utility elicitation," in *Proceedings of National Conference on Artificial Intelligence*, 2000, pp. 363–369.

[17] T. Wang and C. Boutilier, "Incremental utility elicitation with the minimax regret decision criterion," in *Proceedings of IJCAI*, 2003, pp. 309–318.

[18] H. Zhang and D. C. Parkes, "Value-based policy teaching with active indirect elicitation," in *Proceedings of AAAI*, vol. 8, 2008, pp. 208–214.

[19] A. Y. Ng, S. J. Russell, *et al.*, "Algorithms for inverse reinforcement learning," in *Proceedings of ICML*, 2000, pp. 663–670.

[20] A. Sciutti, L. Patane, F. Nori, and G. Sandini, "Understanding object weight from human and humanoid lifting actions," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 2, pp. 80–92, 2014.

[21] A. Zhou, D. Hadfield-Menell, A. Nagabandi, and A. D. Dragan, "Expressive robot motion timing," in *Proceedings of HRI*, 2017.

[22] A. Dragan and S. Srinivasa, "Integrating human observer inferences into robot motion planning," *Autonomous Robots*, vol. 37, no. 4, pp. 351–368, 2014.

[23] J. F. Fisac, C. Liu, J. B. Hamrick, S. Sastry, J. K. Hedrick, T. L. Griffiths, and A. D. Dragan, "Generating plans that predict themselves," in *Proceedings of WAFR*, 2016.

[24] E. Cha, A. D. Dragan, and S. S. Srinivasa, "Perceived robot capability," in *Proceedings of RO-MAN*. IEEE, 2015, pp. 541–548.

[25] D. Szafir, B. Mutlu, and T. Fong, "Communication of intent in assistive free flyers," in *Proceedings of HRI*. ACM/IEEE, 2014, pp. 358–365.

[26] Y. Zhang, S. Sreedharan, A. Kulkarni, T. Chakraborti, H. H. Zhuo, and S. Kambhampati, "Plan explicability for robot task planning," in *Proceedings of the RSS Workshop on Planning for Human-Robot Interaction: Shared Autonomy and Collaborative Robotics*, 2016.

[27] A. Powers and S. Kiesler, "The advisor robot: tracing people's mental model from a robot's physical attributes," in *Proceedings of HRI*. ACM/IEEE, 2006, pp. 218–225.

[28] M. J. Gielniak, C. K. Liu, and A. L. Thomaz, "Generating human-like motion for robots," *IJRR*, vol. 32, no. 11, pp. 1275–1301, 2013.

[29] M. N. Nicolescu and M. J. Mataric, "Natural methods for robot task learning: Instructive demonstrations, generalization and practice," in *Proceedings of AAMAS*. ACM, 2003, pp. 241–248.

[30] G. Farin, "Class a bezier curves," *Computer Aided Geometric Design*, vol. 23, no. 7, pp. 573–581, 2006.

[31] ——, *Curves and surfaces for computer-aided geometric design: a practical guide*. Elsevier, 2014.

[32] K. Perlin, "An image synthesizer," *SIGGRAPH Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985.

[33] K. Perlin and E. M. Hoffert, "Hypertexture," in *SIGGRAPH Computer Graphics*, vol. 23, no. 3. ACM, 1989, pp. 253–262.

[34] K. Perlin, "Improving noise," in *Transactions on Graphics*, vol. 21, no. 3. ACM, 2002, pp. 681–682.

IX. Appendix

	Optimality Ratio		Preference Range		Subjective Confidence	
	0 Comp	4 Comp	0 Comp	4 Comp	0 Comp	4 Comp
0 Infl	0.62(0.24)	0.51(0.18)	1530(391)	1271(324)	2.81(0.97)	3.17(1.07)
2 Infl	0.81(0.18)	0.62(0.21)	1860(265)	1544(298)	3.27(0.92)	3.56(0.96)
4 Infl	0.80(0.19)	0.72(0.19)	1861(286)	1626(304)	3.34(1.04)	3.57(0.89)

TABLE I: Mean (std. dev.) optimality ratio, preference range, and subjective confidence for preference maps

	0 Compromise		4 Compromise	
	Same	Different	Same	Different
0 Inflection	0.99 (0.04)	0.99 (0.04)	0.95 (0.08)	0.95 (0.08)
2 Inflection	0.95 (0.06)	0.95 (0.07)	0.91 (0.10)	0.87 (0.10)
4 Inflection	0.95 (0.07)	0.93 (0.08)	0.86 (0.11)	0.83 (0.09)

TABLE II: Mean (std. dev.) optimality ratio for no preference maps

	0 Compromise		4 Compromise	
	Same	Different	Same	Different
0 Inflection	1434 (563)	1434 (563)	1918 (194)	1918 (194)
2 Inflection	1716 (395)	1572 (482)	989 (655)	866 (485)
4 Inflection	1738 (351)	1671 (436)	797 (485)	658 (428)

TABLE III: Mean (std. dev.) preference range for no preference maps

	0 Compromise		4 Compromise	
	Same	Different	Same	Different
0 Inflection	4.24 (0.84)	4.24 (0.84)	3.50 (1.10)	3.50 (1.10)
2 Inflection	3.40 (1.04)	3.59 (0.99)	3.26 (1.13)	3.27 (0.97)
4 Inflection	3.24 (1.13)	3.40 (1.09)	3.37 (1.09)	3.31 (1.06)

TABLE IV: Mean (std. dev.) subjective confidence for no preference maps