This assessment is similar to the "simulation" assessment used by **?**. The metric we consider for the best demonstration is the complement of the normalized regret:

$$BD = 1 - \frac{R(\xi^*) - R(\xi^H)}{R(\xi^*)}$$

$R(\xi^*)$ is the reward for the optimal trajectory ξ^* and $R(\xi^H)$ is the reward for the human's demonstration ξ^H. We normalize regret in this case, because all of the other assessment metrics are normalized, and we take the complement, since larger values indicate better understanding for the other assessments. These two steps allowed us to combine the four assessment metrics into composite metrics more readily.

Composite Metrics

Finally, we propose three composite metrics based on combinations of four individual metrics. Just as we divided reward explanation techniques into feature space techniques and policy space techniques, we can similarly divide our assessment metrics into feature space metrics, which ask directly about features and their weights (FR and FS), and policy space metrics, which ask about the behaviors that result from reward functions (PE and BD). All four individual metrics are normalized, and so we weight them equally in combining them to form the composite metrics. Accordingly, we propose the composite feature space metric: $F = FR + FS$ We also propose the composite policy space metric: $P = PE + BD$ Finally, we propose the overall composite metric: $C = F + P$

Axes of Domain Complexity

In characterizing domains, we consider four different axes of complexity: reward function complexity, feature complexity, environment complexity, and situational complexity. When considering reward functions that are linear in features, we can vary reward function complexity by considering reward functions with more or fewer features. Feature complexity is related to how complex each individual feature within the linear reward function is. While there might be many ways to characterize feature complexity, we consider the interpretability of individual features as a measure of their complexity. Environment complexity includes factors such as the size of the state and action spaces, whether the state and action spaces are discrete or continuous, or whether a domain is Markovian or non-Markovian. Finally, situational complexity indicates whether a person will need to perform other tasks at the same time as receiving the explanation and the number and difficulty of those tasks.

Proposed Experiment

We propose an experiment to test each of the different explanation techniques in domains of varying complexity as defined by the four axes outlined in the previous section. We have selected four domains in order to cover a broad range of complexities. Our domains include a simple grid world scenario, OpenAI Gym's Lunar Lander (**?**), the threats and waypoints domain proposed by **?**, and the threats and waypoints domain combined with a secondary task in which the human needs to monitor a robot traversing in rocky terrain.

Hypotheses

Our hypotheses for the proposed experiment include those listed below. We intend to assess our hypotheses using the proposed metrics for reward understanding.

Hypothesis 1 *Feature space techniques will lead to better reward understanding than policy space techniques in domains of low versus high reward, feature, and environment complexity.*

Hypothesis 2 *Policy space techniques will lead to better reward understanding than feature space techniques in domains of high versus low reward, feature, and environment complexity.*

Hypothesis 3 *The best modality of information (feature versus policy space) will not change between scenarios with low versus high situational complexity in domains of the same reward, feature, and environment complexities.*

Hypothesis 4 *Reward understanding will be worse in scenarios with high versus low situational complexity for both feature space techniques and policy space techniques.*

Conclusion

In this paper, we define categories of existing reward explanation techniques representing a broad set of explanation modalities, and we identify a specific approach we plan to implement in the context of a human subject experiment from each category. We also suggest a suite of assessment techniques and metrics for human reward understanding. These techniques and metrics integrate multiple modalities of human information understanding, including both feature-based information and behavior-/policy-based information. Finally, we define four axes of domain complexity and outline a future experiment to better understand which reward explanation techniques are most effective in which contexts. We hope that the proposed characterization of reward explanation techniques along with the assessment techniques and metrics will contribute to a more systematic understanding of which reward explanation techniques are most beneficial in different contexts through future human subject experiments. Minima esse illo deserunt praesentium, eum recusandae tenetur assumenda voluptatem quas illo fugit, deleniti saepe aspernatur totam consequuntur cumque error velit, explicabo officiis incidunt rerum ipsum voluptatum expedita, rerum tempora veritatis voluptatum culpa.Molestias voluptates ex tenetur doloremque repellat doloribus aut, deleniti numquam odit adipisci quo nihil.Cupiditate nobis obcaecati est nemo repellat vitae maiores nisi, ipsum culpa ullam expedita eaque blanditiis ratione qui nostrum?Quasi a rem dolorum porro at laboriosam provident, voluptatibus natus vitae accusamus harum laborum laboriosam fugit, atque nostrum odit quas provident voluptatibus nobis nam at modi, corrupti officiis praesentium quod possimus, odit veniam nesciunt perspiciatis?Earum deleniti quaerat dicta nam cupiditate reprehenderit delectus expedita alias, nam minima labore vitae animi quasi obcaecati fugit architecto, esse iure exercitationem debitis?Aut possimus deserunt officia ut nulla

maxime assumenda omnis numquam iusto, mollitia explicabo quisquam nisi provident libero?Accusantium soluta hic illum quos maiores deserunt blanditiis, culpa fugiat vel doloribus ipsa sunt magnam consectetur ipsam