16-886 Embodied AI Safety

Research Skills

Technical Writing

Andrea Bajcsy

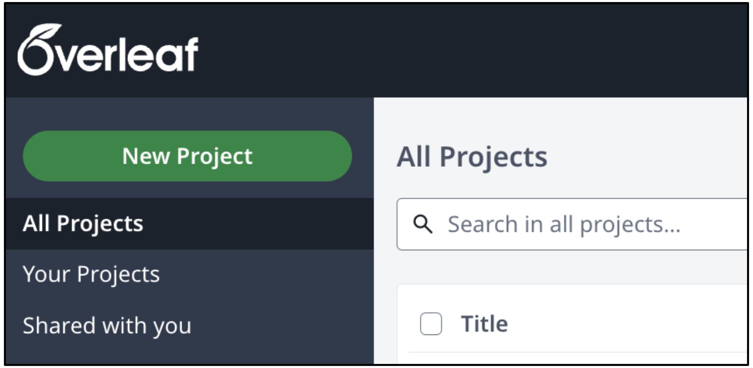
abajcsy@cmu.edu

Carnegie Mellon University

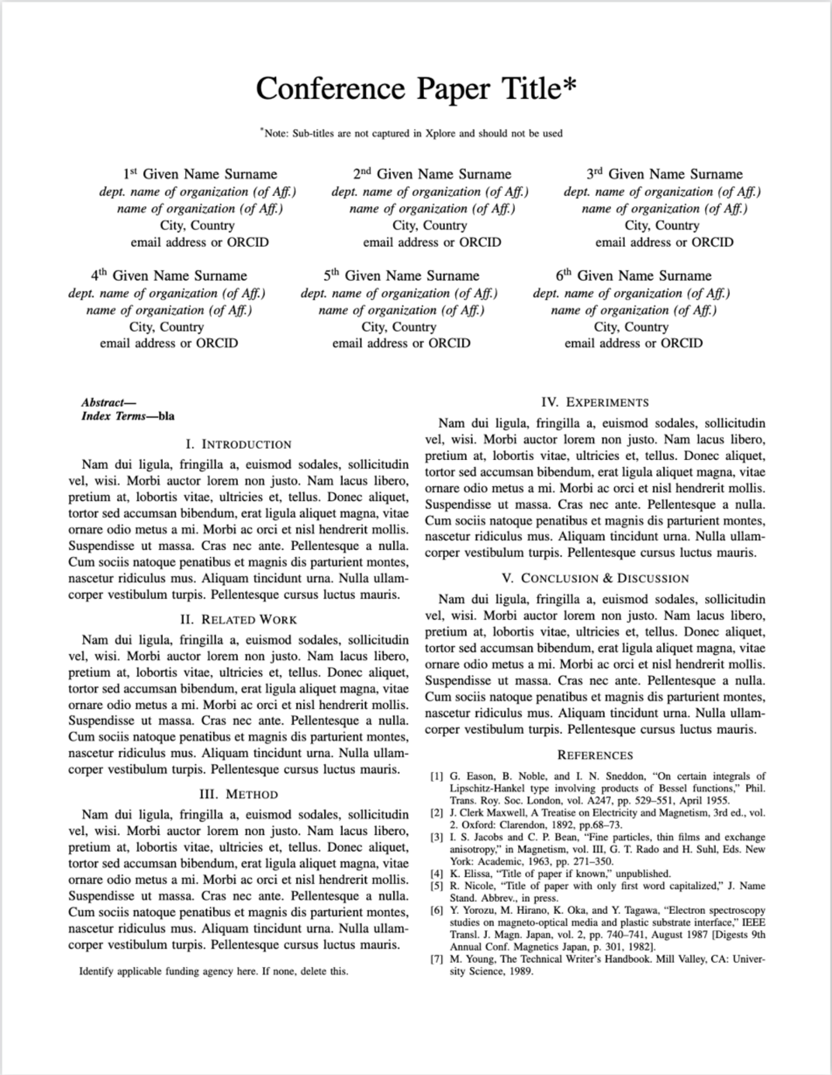
intent ROBOTICS LAB

You have been doing research, you've been generating results, its time to write it up!

Open up Overleaf…



…create a new paper document



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—bla**

I. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

II. Related Work

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

III. Method

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

IV. Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

V. Conclusion & Discussion

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

References

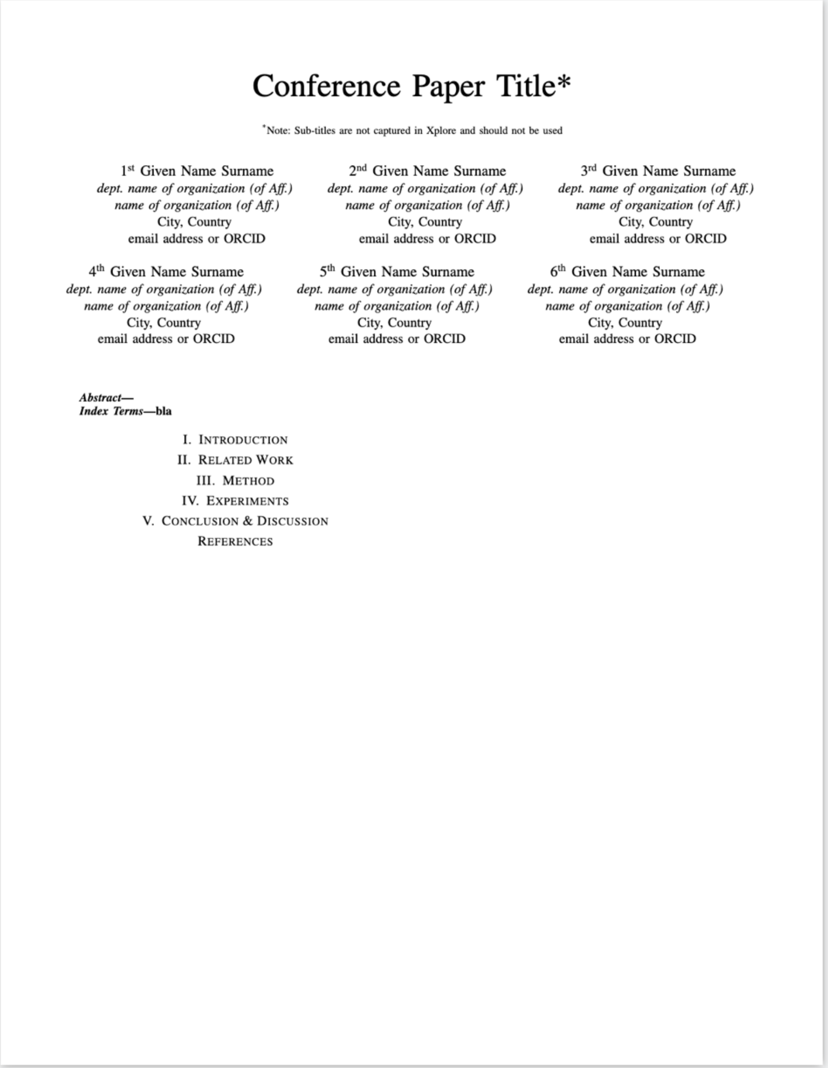
[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, “Title of paper if known,” unpublished.
[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Identify applicable funding agency here. If none, delete this.

More realistically….

😱

How do you start?



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. INTRODUCTION

II. RELATED WORK

III. METHOD

IV. EXPERIMENTS

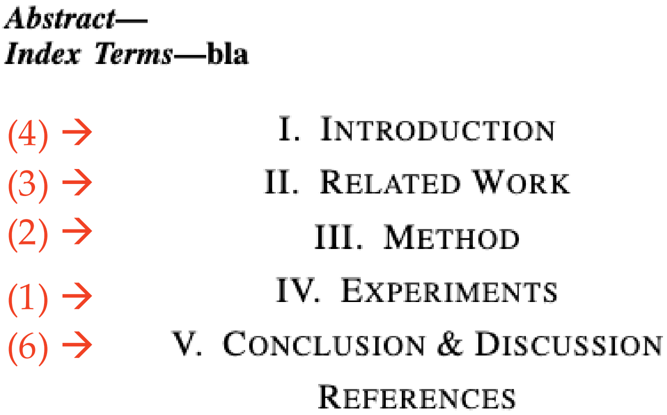
V. CONCLUSION & DISCUSSION

REFERENCES

Order in which I write sections:

Why?

- Most concrete to most abstract
- Experimental results **govern the story I can tell** from the start



(5) → ***Abstract*—**
***Index Terms*—bla**

(4) → I. INTRODUCTION

(3) → II. RELATED WORK

(2) → III. METHOD

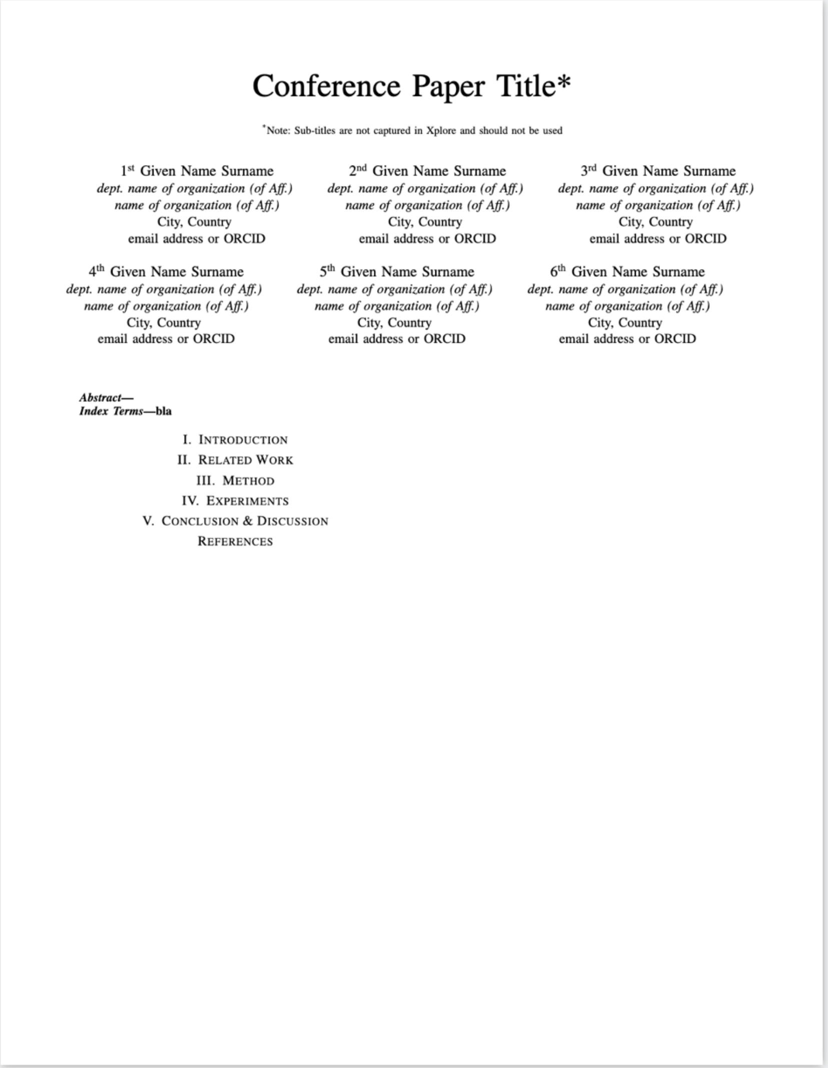
(1) → IV. EXPERIMENTS

(6) → V. CONCLUSION & DISCUSSION

REFERENCES

😱

Ok but how do you ***really*** *start?*



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

V. Conclusion & Discussion

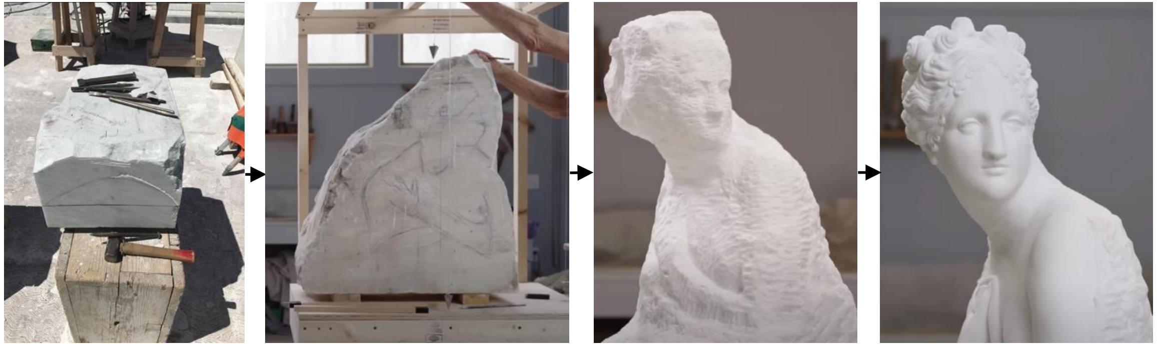
References

Analogy

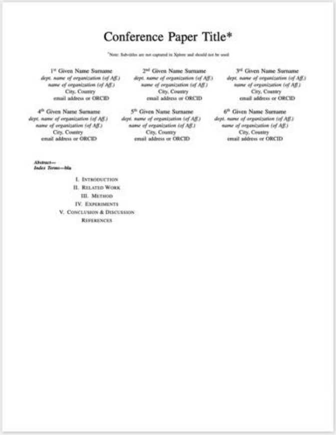
🔑 *Don't get bogged down by the details*

coarse → *fine*

Work Coarse to Fine-Grained



coarse → *fine*



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

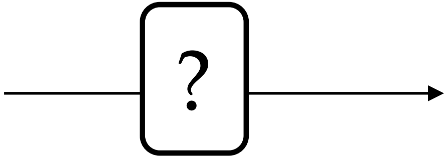
4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

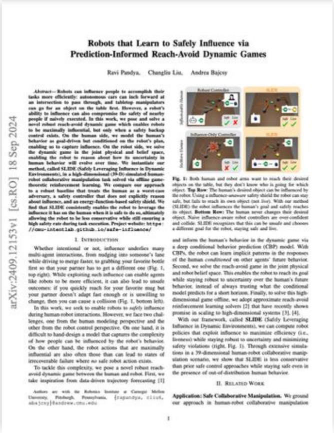
5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction
II. Related Work
III. Method
IV. Experiments
V. Conclusion & Discussion
References





arXiv:2409.12153v1 [cs.RO] 18 Sep 2024

Robots that Learn to Safely Influence via Prediction-Informed Reach-Avoid Dynamic Games

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row**: The human's desired object can be influenced by the robot. Using a influence-unaware safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row:** The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be *unsafe* and chooses a different goal for the robot, staying safe and live.

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from nudging into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where *no* safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' future behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

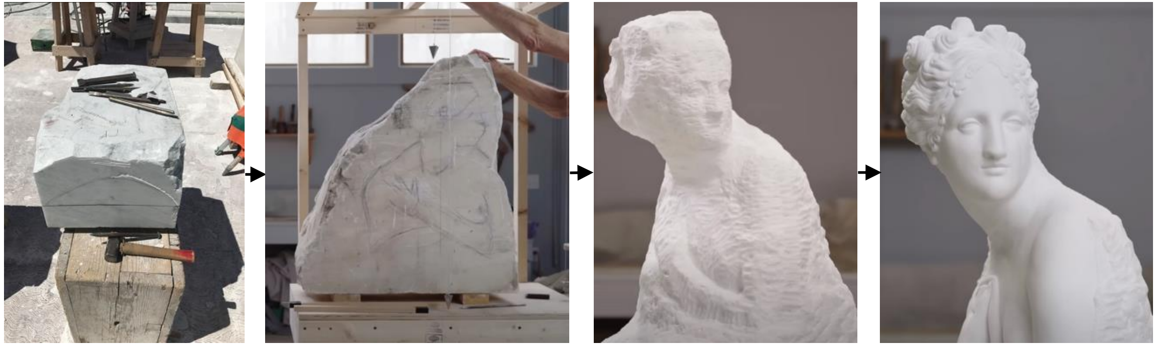
II. Related Work

Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. `{rapandya, cliu6, abajcsy}@andrew.cmu.edu`

1

Work Coarse to Fine-Grained



coarse → *fine*

Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

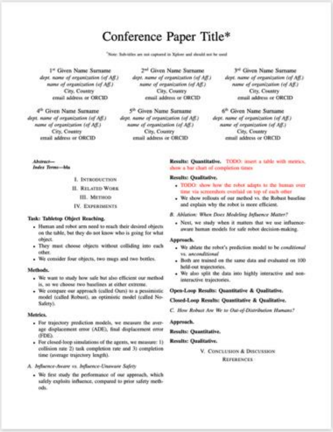
II. Related Work

III. Method

IV. Experiments

V. Conclusion & Discussion

References



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional* vs. *unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

Approach.

Results: Quantitative.

Results: Qualitative.

V. Conclusion & Discussion

References

arXiv:2409.12153v1 [cs.RO] 18 Sep 2024

Robots that Learn to Safely Influence via Prediction-Informed Reach-Avoid Dynamic Games

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row**: The human's desired object can be influenced by the robot. Using a influence-unaware safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row:** The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be unsafe and chooses a different goal for the robot, staying safe and live.

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from nudging into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top-right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where no safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' future behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

II. Related Work

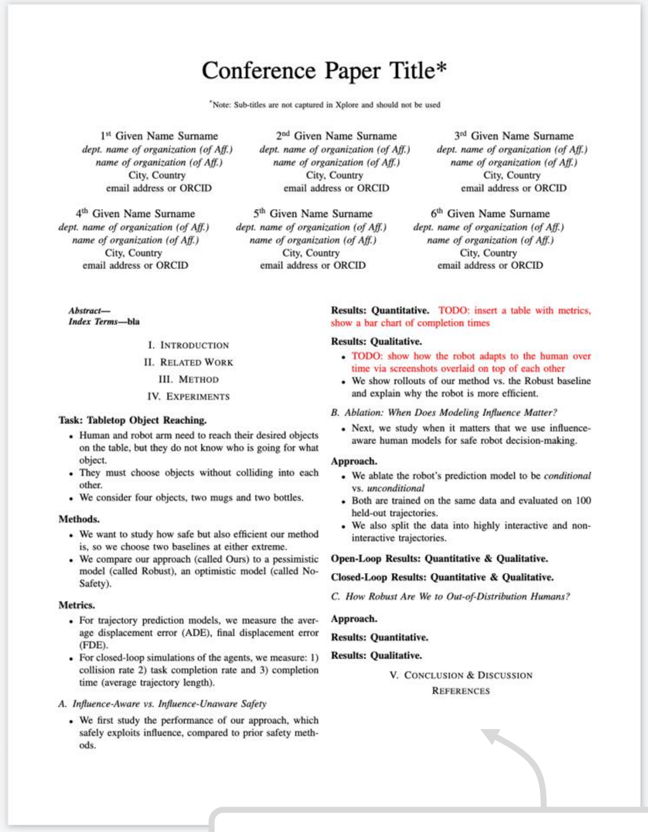
Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania, {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1

Coarse: Minimum Necessary Sketch

If you removed even one of these lines, the sketch wouldn't make sense!



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional* vs. *unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

Approach.

Results: Quantitative.

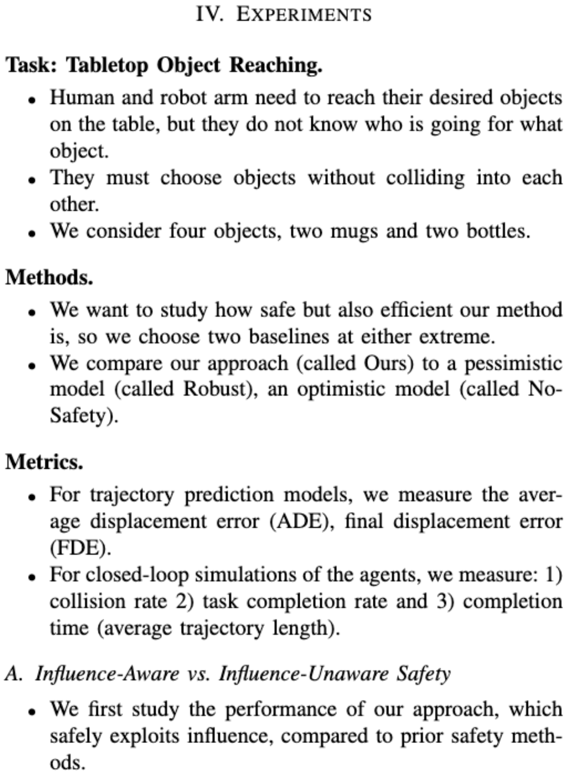
Results: Qualitative.

V. Conclusion & Discussion

References

Same goes for our next level of writing

Coarse: Minimum Necessary Sketch



IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

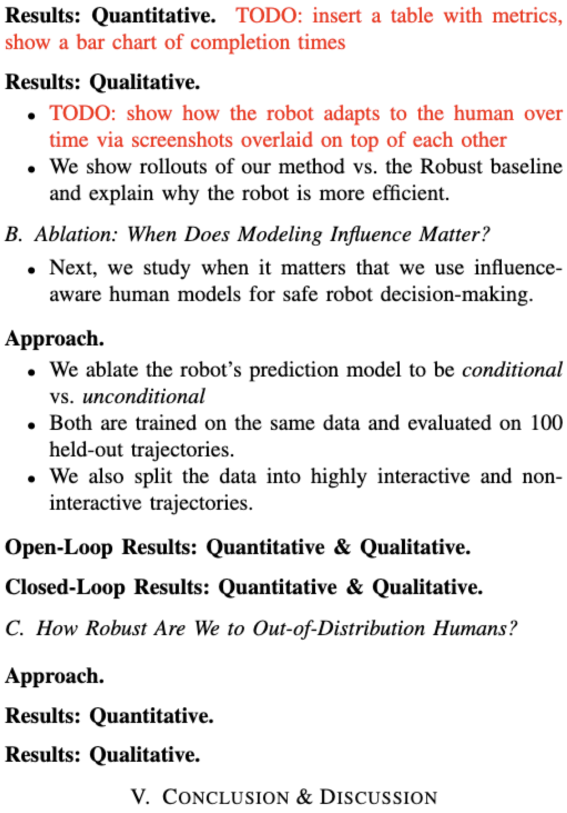
Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Coarse: Minimum Necessary Sketch



Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional* vs. *unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

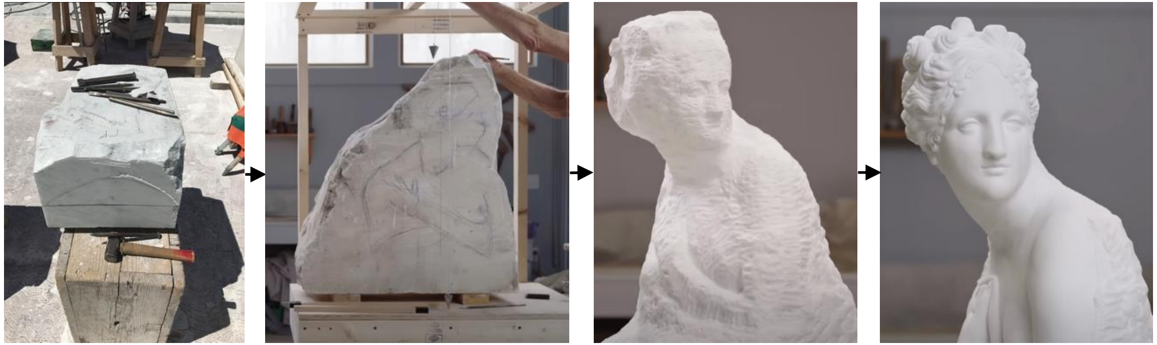
Approach.

Results: Quantitative.

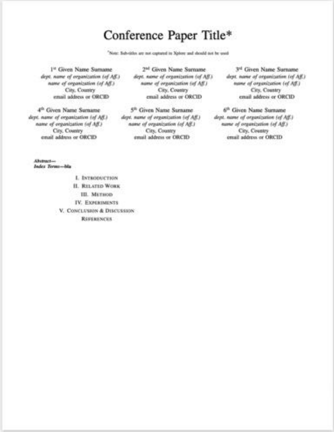
Results: Qualitative.

V. Conclusion & Discussion

Work Coarse to Fine-Grained



coarse → *fine*



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

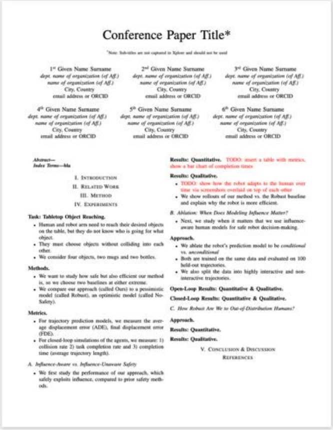
II. Related Work

III. Method

IV. Experiments

V. Conclusion & Discussion

References



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional* vs. *unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

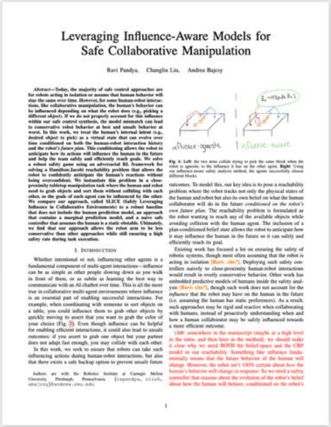
Approach.

Results: Quantitative.

Results: Qualitative.

V. Conclusion & Discussion

References



Leveraging Influence-Aware Models for Safe Collaborative Manipulation

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking a different object). If we do not properly account for this influence within our safe control synthesis, the model mismatch can lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired object to pick) as a virtual state that can evolve over time conditioned on both the human-robot interaction history and the robot's *future plan*. This conditioning allows the robot to anticipate how its actions will influence the human in the future and help the team safely and efficiently reach goals. We solve a robust safety game using an adversarial RL framework for solving a Hamilton-Jacobi reachability problem that allows the robot to confidently anticipate the human's reactions without being overconfident. We instantiate this problem in a close-proximity tabletop manipulation task where the human and robot need to grab objects and sort them without colliding with each other, so the goals of each agent can be influenced by the other. We compare our approach, called SLICE (Safely Leveraging Influence in Collaborative Environments) to a robust baseline that does not include the human prediction model, an approach that contains a marginal prediction model, and a naive safe controller that assumes the human is a static obstacle. Ultimately, we find that our approach allows the robot arm to be less conservative than other approaches while still ensuring a high safety rate during task execution.**

Fig. 1: **Left**: the two arms collide trying to pick the same block when the robot is agnostic to the influence it has on the other agent. **Right**: Using our influence-aware safety analysis method, the agents successfully choose different blocks.

I. Introduction

Whether intentional or not, influencing other agents is a fundamental component of multi-agent interactions—influence can be as simple as other people slowing down as you walk in front of them, or as subtle as learning the best way to communicate with an AI chatbot over time. This is all the more true in collaborative multi-agent environments where influence is an essential part of enabling successful interactions. For example, when coordinating with someone to sort objects on a table, you could influence them to grab other objects by quickly moving to assert that you want to grab the color of your choice (Fig. 1). Even though influence can be helpful for enabling efficient interactions, it could also lead to unsafe outcomes: if you assert to grab one object but your partner does not adapt fast enough, you may collide with each other.

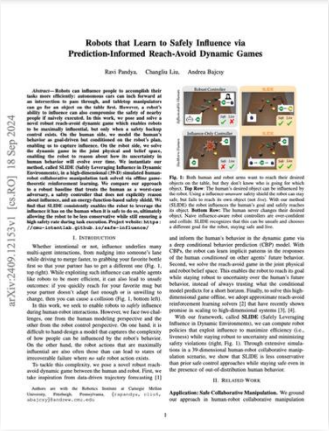
In this work, we seek to ensure that robots can take such influencing actions during human-robot interactions, but also that there exists a safe backup option to prevent unsafe future outcomes. To model this, our key idea is to pose a reachability problem where the robot tracks not only the physical states of the human and robot but also its own belief on what the human collaborator will do in the future *conditioned on the robot's own future plan*. The reachability problem is formulated as the robot wanting to reach any of the available objects while avoiding collision with the human agent. The inclusion of a plan-conditioned belief state allows the robot to anticipate how it may influence the human in the future so it can safely and efficiently reach its goal.

Existing work has focused a lot on ensuring the safety of robotic systems, though most often assuming that the robot is acting in isolation [Ravi: cite?]. Deploying such safety controllers naively to close-proximity human-robot interactions would result in overly conservative behavior. Other work has embedded predictive models of humans inside the safety analysis [Ravi: cite?], though such work does not account for the *influence* that the robot may have on the human in the future (i.e. assuming the human has static preferences). As a result, such approaches may be rigid and reactive when collaborating with humans, instead of proactively understanding when and how a human collaborator may be safely influenced towards a more efficient outcome.

(AB: somewhere in the manuscript (maybe at a high level in the intro, and then later in the method), we should make it clear why we need BOTH the belief-space and the CBP model in our reachability. Something like influence fundamentally means that the future behavior of the human will change. However, the robot isn't 100% certain about *how* the human's behavior will change in response. So we need a safety controller that reasons about the evolution of the robot's belief about how the human will behave, conditioned on the robot's

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1



arXiv:2409.12153v1 [cs.RO] 18 Sep 2024

Robots that Learn to Safely Influence via Prediction-Informed Reach-Avoid Dynamic Games

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row**: The human's desired object can be influenced by the robot. Using a influence-unaware safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row**: The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be unsafe and chooses a different goal for the robot, staying safe and live.

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from nudging into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where *no* safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' future behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

II. Related Work

Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1

Finer: Flesh Out the Paper

Only after the sketch, are we ready to flesh out the sculpture's key features

This carving on the left is already looking like a sculpture!

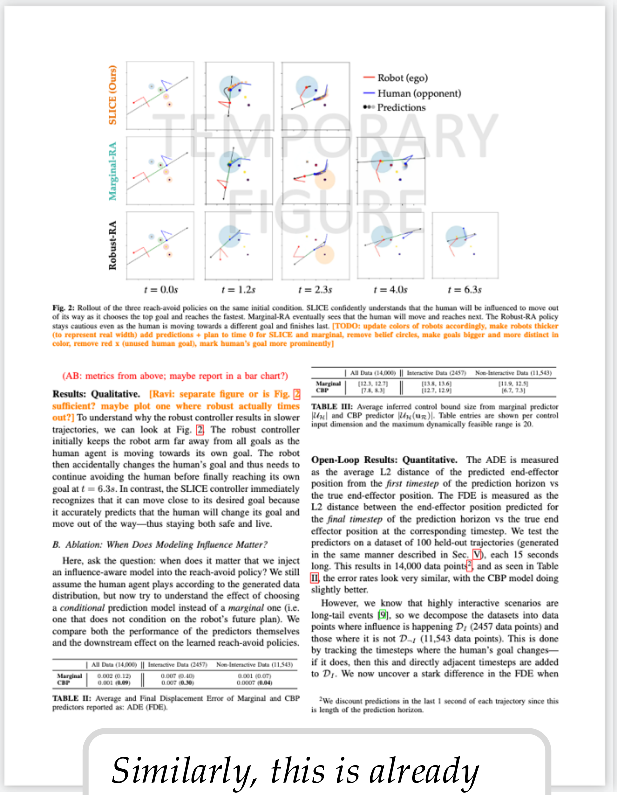


Fig. 2: Rollout of the three reach-avoid policies on the same initial condition. SLICE confidently understands that the human will be influenced to move out of its way as it chooses the top goal and reaches the fastest. Marginal-RA eventually sees that the human will move and reaches next. The Robust-RA policy stays cautious even as the human is moving towards a different goal and finishes last. **[TODO: update colors of robots accordingly, make robots thicker (to represent real width) add predictions + plan to time 0 for SLICE and marginal, remove belief circles, make goals bigger and more distinct in color, remove red x (unused human goal), mark human's goal more prominently]**

(AB: metrics from above; maybe report in a bar chart?)

Results: Qualitative. **[Ravi: separate figure or is Fig. 2 sufficient? maybe plot one where robust actually times out?]** To understand why the robust controller results in slower trajectories, we can look at Fig. 2. The robust controller initially keeps the robot arm far away from all goals as the human agent is moving towards its own goal. The robot then accidentally changes the human's goal and thus needs to continue avoiding the human before finally reaching its own goal at $t = 6.3s$. In contrast, the SLICE controller immediately recognizes that it can move close to its desired goal because it accurately predicts that the human will change its goal and move out of the way—thus staying both safe and live.

B. Ablation: When Does Modeling Influence Matter?

Here, ask the question: when does it matter that we inject an influence-aware model into the reach-avoid policy? We still assume the human agent plays according to the generated data distribution, but now try to understand the effect of choosing a *conditional* prediction model instead of a *marginal* one (i.e. one that does not condition on the robot's future plan). We compare both the performance of the predictors themselves and the downstream effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Average and Final Displacement Error of Marginal and CBP predictors reported as: ADE (FDE).

	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	[12.3, 12.7]	[13.8, 13.6]	[11.9, 12.5]
CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

TABLE III: Average inferred control bound size from marginal predictor $|\mathcal{U}_{\mathcal{H}}|$ and CBP predictor $|\mathcal{U}_{\mathcal{H}}(\mathbf{u}_{\mathcal{R}})|$. Table entries are shown per control input dimension and the maximum dynamically feasible range is 20.

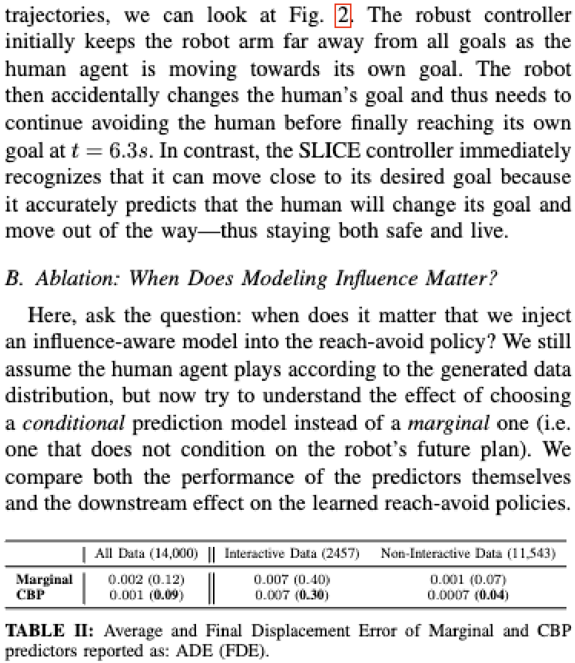
Open-Loop Results: Quantitative. The ADE is measured as the average L2 distance of the predicted end-effector position from the *first timestep* of the prediction horizon vs the true end-effector position. The FDE is measured as the L2 distance between the end-effector position predicted for the *final timestep* of the prediction horizon vs the true end effector position at the corresponding timestep. We test the predictors on a dataset of 100 held-out trajectories (generated in the same manner described in Sec. V), each 15 seconds long. This results in 14,000 data points[2], and as seen in Table II, the error rates look very similar, with the CBP model doing slightly better.

However, we know that highly interactive scenarios are long-tail events [9], so we decompose the datasets into data points where influence is happening $\mathcal{D}_I$ (2457 data points) and those where it is not $\mathcal{D}_{\neg I}$ (11,543 data points). This is done by tracking the timesteps where the human's goal changes—if it does, then this and directly adjacent timesteps are added to $\mathcal{D}_I$. We now uncover a stark difference in the FDE when

[2] We discount predictions in the last 1 second of each trajectory since this is length of the prediction horizon.

Similarly, this is already looking like a paper!

Finer: Flesh Out the Paper



trajectories, we can look at Fig. 2. The robust controller initially keeps the robot arm far away from all goals as the human agent is moving towards its own goal. The robot then accidentally changes the human's goal and thus needs to continue avoiding the human before finally reaching its own goal at $t = 6.3s$. In contrast, the SLICE controller immediately recognizes that it can move close to its desired goal because it accurately predicts that the human will change its goal and move out of the way—thus staying both safe and live.

B. Ablation: When Does Modeling Influence Matter?

Here, ask the question: when does it matter that we inject an influence-aware model into the reach-avoid policy? We still assume the human agent plays according to the generated data distribution, but now try to understand the effect of choosing a *conditional* prediction model instead of a *marginal* one (i.e. one that does not condition on the robot's future plan). We compare both the performance of the predictors themselves and the downstream effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Average and Final Displacement Error of Marginal and CBP predictors reported as: ADE (FDE).

This is a good time to add in results

Good time to convert bullet pts into sentences

Finer: Flesh Out the Paper

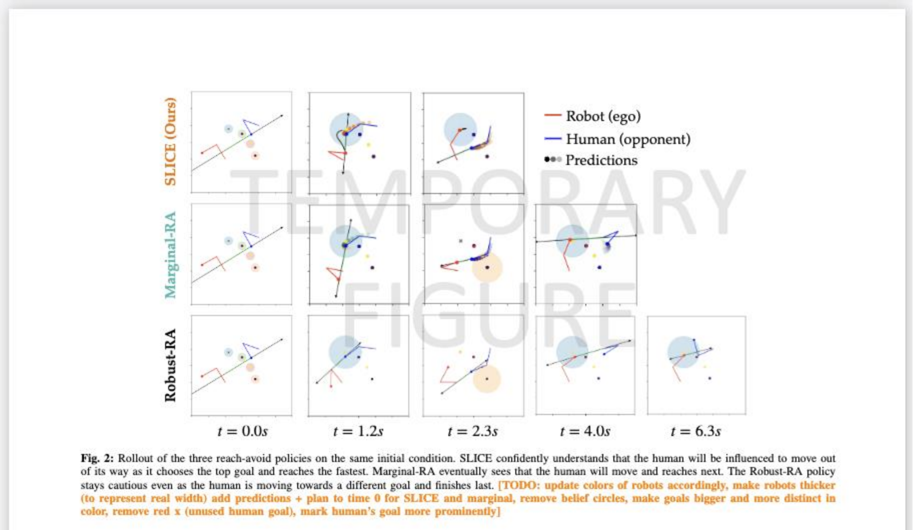
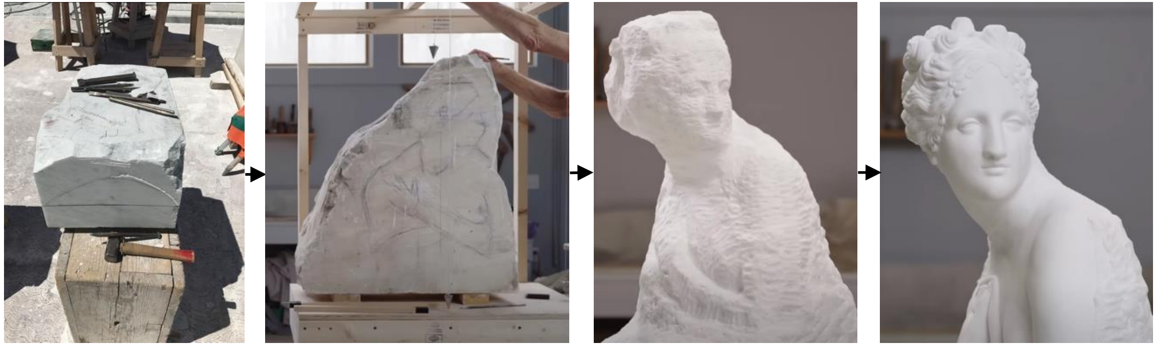


Fig. 2: Rollout of the three reach-avoid policies on the same initial condition. SLICE confidently understands that the human will be influenced to move out of its way as it chooses the top goal and reaches the fastest. Marginal-RA eventually sees that the human will move and reaches next. The Robust-RA policy stays cautious even as the human is moving towards a different goal and finishes last. **[TODO: update colors of robots accordingly, make robots thicker (to represent real width) add predictions + plan to time 0 for SLICE and marginal, remove belief circles, make goals bigger and more distinct in color, remove red x (unused human goal), mark human's goal more prominently]**

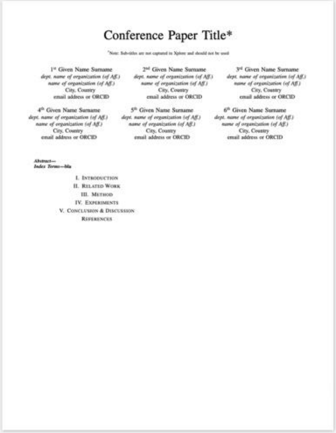
This is the right stage to think:

- What additional results / visuals would help me make my claims clearer?
- Do I need to revisit the way I chose to organize my content and arguments?

Work Coarse to Fine-Grained



coarse → *fine*



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

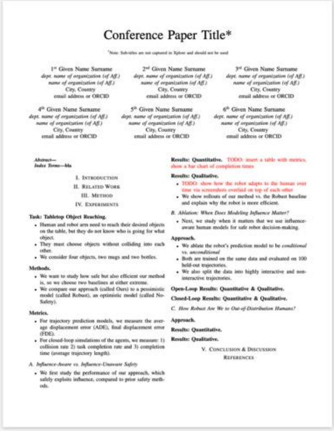
II. Related Work

III. Method

IV. Experiments

V. Conclusion & Discussion

References



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional* vs. *unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

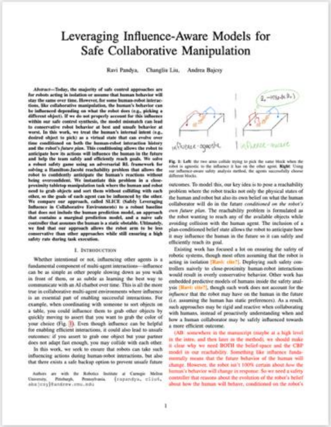
Approach.

Results: Quantitative.

Results: Qualitative.

V. Conclusion & Discussion

References



Leveraging Influence-Aware Models for
Safe Collaborative Manipulation

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking a different object). If we do not properly account for this influence within our safe control synthesis, the model mismatch can lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired object to pick) as a virtual state that can evolve over time conditioned on both the human-robot interaction history and the robot's *future plan*. This conditioning allows the robot to anticipate how its actions will influence the human in the future and help the team safely and efficiently reach goals. We solve a robust safety game using an adversarial RL framework for solving a Hamilton-Jacobi reachability problem that allows the robot to confidently anticipate the human's reactions without being overconfident. We instantiate this problem in a close-proximity tabletop manipulation task where the human and robot need to grab objects and sort them without colliding with each other, so the goals of each agent can be influenced by the other. We compare our approach, called SLICE (Safely Leveraging Influence in Collaborative Environments) to a robust baseline that does not include the human prediction model, an approach that contains a marginal prediction model, and a naive safe controller that assumes the human is a static obstacle. Ultimately, we find that our approach allows the robot arm to be less conservative than other approaches while still ensuring a high safety rate during task execution.**

I. Introduction

Whether intentional or not, influencing other agents is a fundamental component of multi-agent interactions—influence can be as simple as other people slowing down as you walk in front of them, or as subtle as learning the best way to communicate with an AI chatbot over time. This is all the more true in collaborative multi-agent environments where influence is an essential part of enabling successful interactions. For example, when coordinating with someone to sort objects on a table, you could influence them to grab other objects by quickly moving to assert that you want to grab the color of your choice (Fig. 1). Even though influence can be helpful for enabling efficient interactions, it could also lead to unsafe outcomes: if you assert to grab one object but your partner does not adapt fast enough, you may collide with each other.

In this work, we seek to ensure that robots can take such influencing actions during human-robot interactions, but also that there exists a safe backup option to prevent unsafe future outcomes. To model this, our key idea is to pose a reachability problem where the robot tracks not only the physical states of the human and robot but also its own belief on what the human collaborator will do in the future *conditioned on the robot's own future plan*. The reachability problem is formulated as the robot wanting to reach any of the available objects while avoiding collision with the human agent. The inclusion of a plan-conditioned belief state allows the robot to anticipate how it may influence the human in the future so it can safely and efficiently reach its goal.

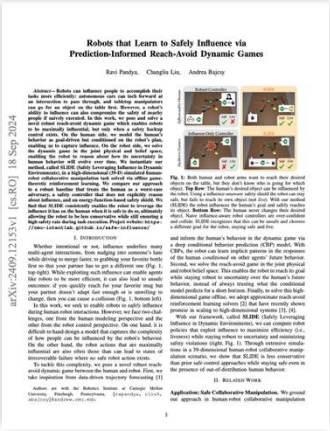
Existing work has focused a lot on ensuring the safety of robotic systems, though most often assuming that the robot is acting in isolation [Ravi: cite?]. Deploying such safety controllers naively to close-proximity human-robot interactions would result in overly conservative behavior. Other work has embedded predictive models of humans inside the safety analysis [Ravi: cite?], though such work does not account for the *influence* that the robot may have on the human in the future (i.e. assuming the human has static preferences). As a result, such approaches may be rigid and reactive when collaborating with humans, instead of proactively understanding when and how a human collaborator may be safely influenced towards a more efficient outcome.

(AB: somewhere in the manuscript (maybe at a high level in the intro, and then later in the method), we should make it clear why we need BOTH the belief-space and the CBF model in our reachability. Something like influence fundamentally means that the future behavior of the human will change. However, the robot isn't 100% certain about *how* the human's behavior will change in response. So we need a safety controller that reasons about the evolution of the robot's belief about how the human will behave, conditioned on the robot's

Fig. 1: Left: the two arms collide trying to pick the same block when the robot is agnostic to the influence it has on the other agent. **Right**: Using our influence-aware safety analysis method, the agents successfully choose different blocks.

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1



arXiv:2409.12153v1 [cs.RO] 18 Sep 2024

**Robots that Learn to Safely Influence via
Prediction-Informed Reach-Avoid Dynamic Games**

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from nudging into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where *no* safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' *future* behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

II. Related Work

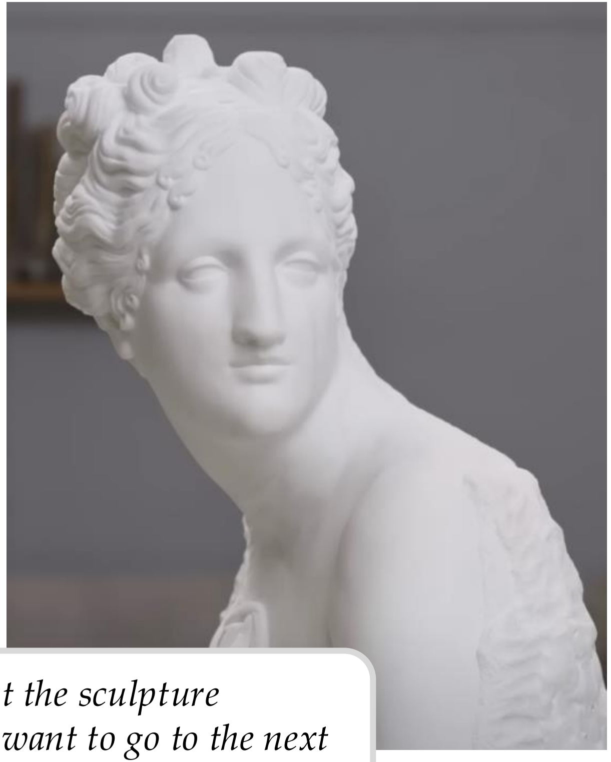
Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row**: The human's desired object can be influenced by the robot. Using a influence-unaware safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row:** The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be unsafe and chooses a different goal for the robot, staying safe and live.

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1

Fine-Grained



You could stop at the sculpture before…but you want to go to the next level! Time to do fine-grained details

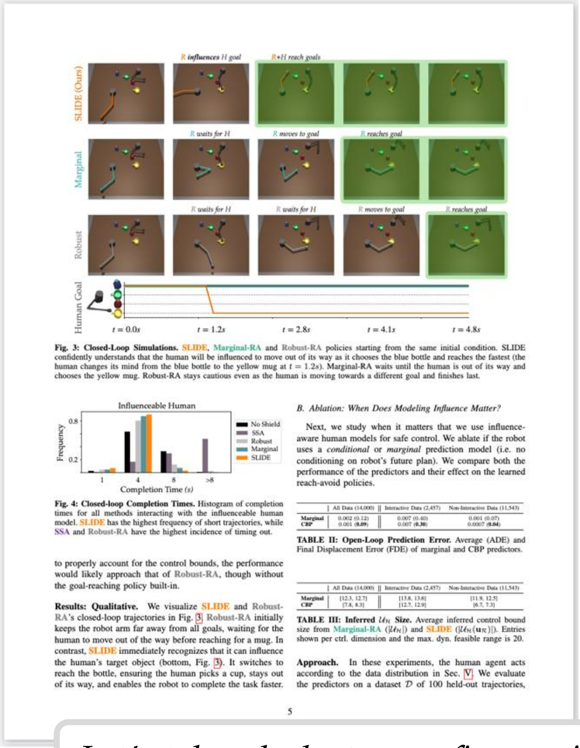


Fig. 3: Closed-Loop Simulations. SLIDE, Marginal-RA and Robust-RA policies starting from the same initial condition. SLIDE confidently understands that the human will be influenced to move out of its way as it chooses the blue bottle and reaches the fastest (the human changes its mind from the blue bottle to the yellow mug at $t = 1.2s$). Marginal-RA waits until the human is out of its way and chooses the yellow mug. Robust-RA stays cautious even as the human is moving towards a different goal and finishes last.

Fig. 4: Closed-loop Completion Times. Histogram of completion times for all methods interacting with the influenceable human model. SLIDE has the highest frequency of short trajectories, while SSA and Robust-RA have the highest incidence of timing out.

to properly account for the control bounds, the performance would likely approach that of Robust-RA, though without the goal-reaching policy built-in.

Results: Qualitative. We visualize SLIDE and Robust-RA's closed-loop trajectories in Fig. 3. Robust-RA initially keeps the robot arm far away from all goals, waiting for the human to move out of the way before reaching for a mug. In contrast, SLIDE immediately recognizes that it can influence the human's target object (bottom, Fig. 3). It switches to reach the bottle, ensuring the human picks a cup, stays out of its way, and enables the robot to complete the task faster.

B. Ablation: When Does Modeling Influence Matter?

Next, we study when it matters that we use influence-aware human models for safe control. We ablate if the robot uses a *conditional* or *marginal* prediction model (i.e. no conditioning on robot's future plan). We compare both the performance of the predictors and their effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Open-Loop Prediction Error. Average (ADE) and Final Displacement Error (FDE) of marginal and CBP predictors.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	[12.3, 12.7]	[13.8, 13.6]	[11.9, 12.5]
CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

TABLE III: Inferred $\mathcal{U}_\mathcal{H}$ Size. Average inferred control bound size from Marginal-RA ($|\mathcal{U}_\mathcal{H}|$) and SLIDE ($|\mathcal{U}_\mathcal{H}(\mathbf{u}_\mathcal{R})|$). Entries shown per ctrl. dimension and the max. dyn. feasible range is 20.

Approach. In these experiments, the human agent acts according to the data distribution in Sec. V. We evaluate the predictors on a dataset $\mathcal{D}$ of 100 held-out trajectories,

Let's take a look at some fine-grained details…

Finer

Fine-Grained

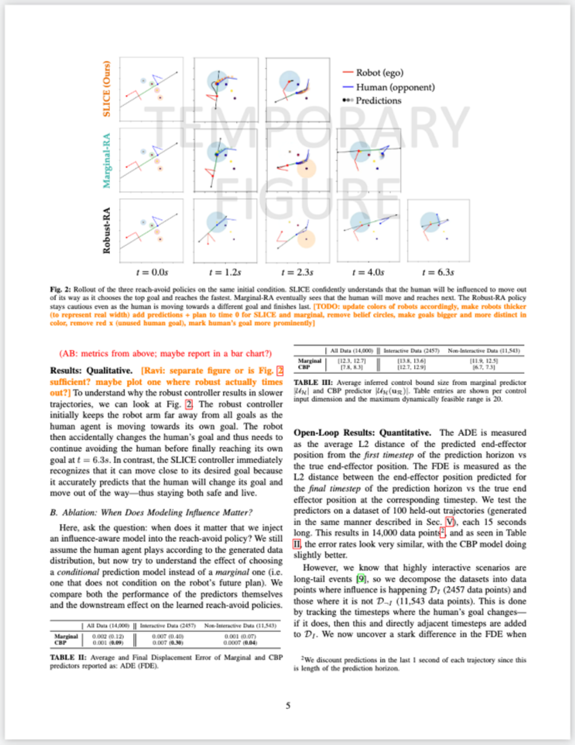


Fig. 2: Rollout of the three reach-avoid policies on the same initial condition. SLICE confidently understands that the human will be influenced to move out of its way as it chooses the top goal and reaches the fastest. Marginal-RA eventually sees that the human will move and reaches next. The Robust-RA policy stays cautious even as the human is moving towards a different goal and finishes last. **[TODO: update colors of robots accordingly, make robots thicker (to represent real width) add predictions + plan to time 0 for SLICE and marginal, remove belief circles, make goals bigger and more distinct in color, remove red x (unused human goal), mark human's goal more prominently]**

(AB: metrics from above; maybe report in a bar chart?)

Results: Qualitative. **[Ravi: separate figure or is Fig. 2 sufficient? maybe plot one where robust actually times out?]** To understand why the robust controller results in slower trajectories, we can look at Fig. 2. The robust controller initially keeps the robot arm far away from all goals as the human agent is moving towards its own goal. The robot then accidentally changes the human's goal and thus needs to continue avoiding the human before finally reaching its own goal at $t = 6.3s$. In contrast, the SLICE controller immediately recognizes that it can move close to its desired goal because it accurately predicts that the human will change its goal and move out of the way—thus staying both safe and live.

B. Ablation: When Does Modeling Influence Matter?

Here, ask the question: when does it matter that we inject an influence-aware model into the reach-avoid policy? We still assume the human agent plays according to the generated data distribution, but now try to understand the effect of choosing a *conditional* prediction model instead of a *marginal* one (i.e. one that does not condition on the robot's future plan). We compare both the performance of the predictors themselves and the downstream effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Average and Final Displacement Error of Marginal and CBP predictors reported as: ADE (FDE).

	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	[12.3, 12.7]	[13.8, 13.6]	[11.9, 12.5]
CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

TABLE III: Average inferred control bound size from marginal predictor $|\mathcal{U}_\mathcal{H}|$ and CBP predictor $|\mathcal{U}_\mathcal{H}(\mathbf{u}_\mathcal{R})|$. Table entries are shown per control input dimension and the maximum dynamically feasible range is 20.

Open-Loop Results: Quantitative. The ADE is measured as the average L2 distance of the predicted end-effector position from the *first timestep* of the prediction horizon vs the true end-effector position. The FDE is measured as the L2 distance between the end-effector position predicted for the *final timestep* of the prediction horizon vs the true end effector position at the corresponding timestep. We test the predictors on a dataset of 100 held-out trajectories (generated in the same manner described in Sec. V), each 15 seconds long. This results in 14,000 data points[2], and as seen in Table II, the error rates look very similar, with the CBP model doing slightly better.

However, we know that highly interactive scenarios are long-tail events [9], so we decompose the datasets into data points where influence is happening $\mathcal{D}_I$ (2457 data points) and those where it is not $\mathcal{D}_{\neg I}$ (11,543 data points). This is done by tracking the timesteps where the human's goal changes—if it does, then this and directly adjacent timesteps are added to $\mathcal{D}_I$. We now uncover a stark difference in the FDE when

[2] We discount predictions in the last 1 second of each trajectory since this is length of the prediction horizon.

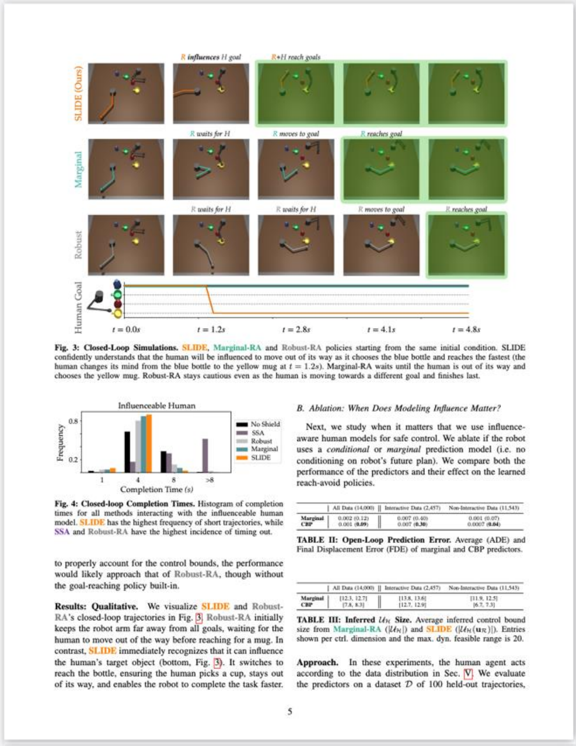


Fig. 3: Closed-Loop Simulations. SLIDE, Marginal-RA and Robust-RA policies starting from the same initial condition. SLIDE confidently understands that the human will be influenced to move out of its way as it chooses the blue bottle and reaches the fastest (the human changes its mind from the blue bottle to the yellow mug at $t = 1.2s$). Marginal-RA waits until the human is out of its way and chooses the yellow mug. Robust-RA stays cautious even as the human is moving towards a different goal and finishes last.

Fig. 4: Closed-loop Completion Times. Histogram of completion times for all methods interacting with the influenceable human model. SLIDE has the highest frequency of short trajectories, while SSA and Robust-RA have the highest incidence of timing out.

to properly account for the control bounds, the performance would likely approach that of Robust-RA, though without the goal-reaching policy built-in.

Results: Qualitative. We visualize SLIDE and Robust-RA's closed-loop trajectories in Fig. 3. Robust-RA initially keeps the robot arm far away from all goals, waiting for the human to move out of the way before reaching for a mug. In contrast, SLIDE immediately recognizes that it can influence the human's target object (bottom, Fig. 3). It switches to reach the bottle, ensuring the human picks a cup, stays out of its way, and enables the robot to complete the task faster.

B. Ablation: When Does Modeling Influence Matter?

Next, we study when it matters that we use influence-aware human models for safe control. We ablate if the robot uses a *conditional* or *marginal* prediction model (i.e. no conditioning on robot's future plan). We compare both the performance of the predictors and their effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Open-Loop Prediction Error. Average (ADE) and Final Displacement Error (FDE) of marginal and CBP predictors.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	[12.3, 12.7]	[13.8, 13.6]	[11.9, 12.5]
CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

TABLE III: Inferred $\mathcal{U}_\mathcal{H}$ Size. Average inferred control bound size from Marginal-RA ($|\mathcal{U}_\mathcal{H}|$) and SLIDE ($|\mathcal{U}_\mathcal{H}(\mathbf{u}_\mathcal{R})|$). Entries shown per ctrl. dimension and the max. dyn. feasible range is 20.

Approach. In these experiments, the human agent acts according to the data distribution in Sec. V. We evaluate the predictors on a dataset $\mathcal{D}$ of 100 held-out trajectories,

Finer

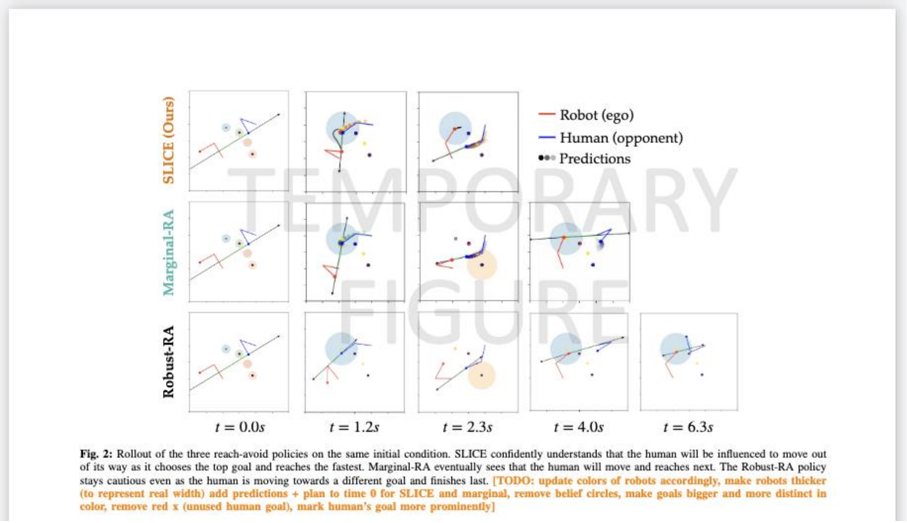


Fig. 2: Rollout of the three reach-avoid policies on the same initial condition. SLICE confidently understands that the human will be influenced to move out of its way as it chooses the top goal and reaches the fastest. Marginal-RA eventually sees that the human will move and reaches next. The Robust-RA policy stays cautious even as the human is moving towards a different goal and finishes last. **[TODO: update colors of robots accordingly, make robots thicker (to represent real width) add predictions + plan to time 0 for SLICE and marginal, remove belief circles, make goals bigger and more distinct in color, remove red x (unused human goal), mark human's goal more prominently]**

Fine-Grained

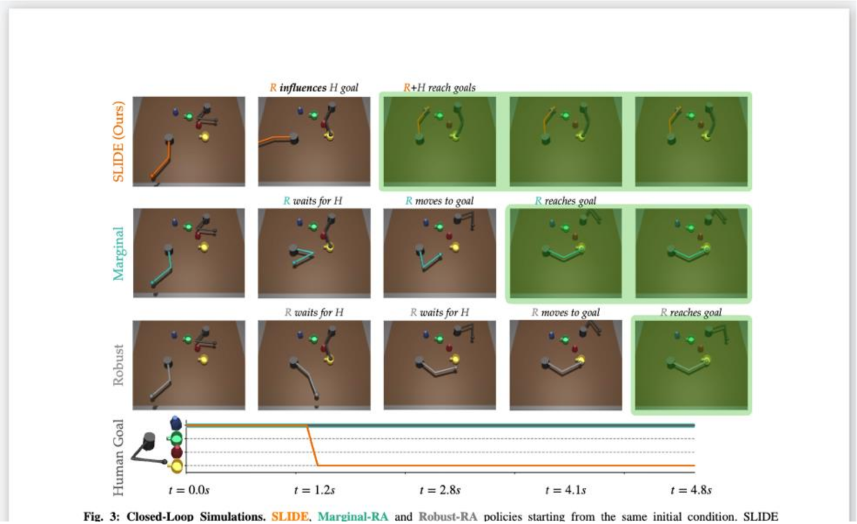


Fig. 3: Closed-Loop Simulations. SLIDE, Marginal-RA and Robust-RA policies starting from the same initial condition. SLIDE

Better rendering

Text annotations

Color coding

Additional (bottom) plot to emphasize key aspect

Finer

(AB: metrics from above; maybe report in a bar chart?)

Results: Qualitative. [Ravi: separate figure or is Fig. 2

Fine-Grained

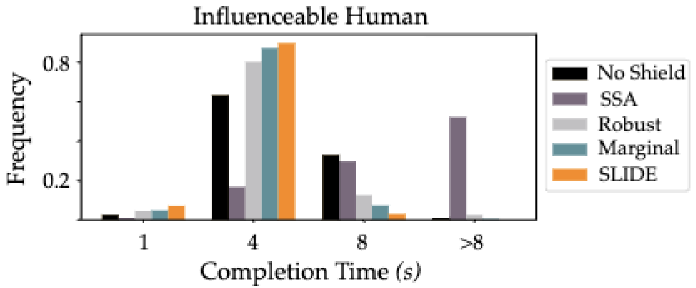
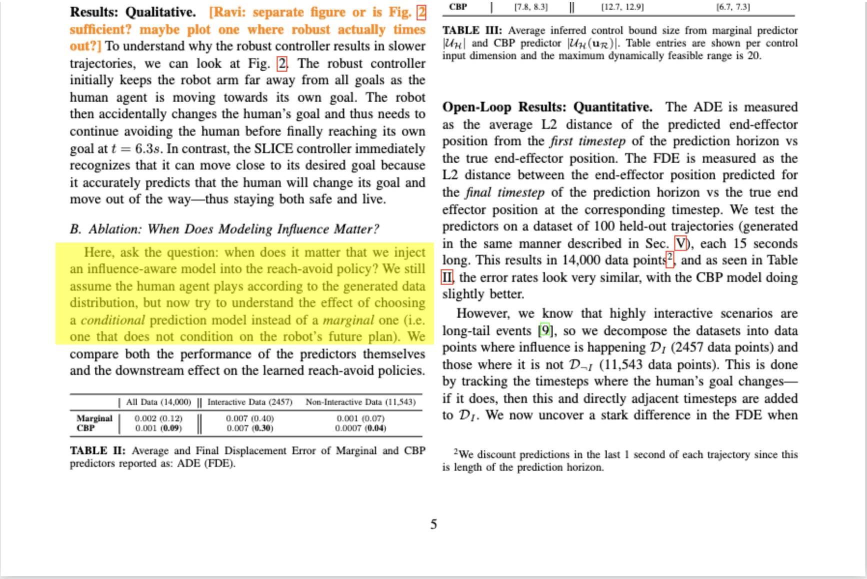


Fig. 4: Closed-loop Completion Times. Histogram of completion times for all methods interacting with the influenceable human model. SLIDE has the highest frequency of short trajectories, while SSA and Robust-RA have the highest incidence of timing out.

Experiment with best type of visualization to make your narrative point

Finer

Fine-Grained



CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

TABLE III: Average inferred control bound size from marginal predictor $|\mathcal{U}_{\mathcal{H}}|$ and CBP predictor $|\mathcal{U}_{\mathcal{H}}(\mathbf{u}_{\mathcal{R}})|$. Table entries are shown per control input dimension and the maximum dynamically feasible range is 20.

Results: Qualitative. **[Ravi: separate figure or is Fig. 2 sufficient? maybe plot one where robust actually times out?]** To understand why the robust controller results in slower trajectories, we can look at Fig. 2. The robust controller initially keeps the robot arm far away from all goals as the human agent is moving towards its own goal. The robot then accidentally changes the human's goal and thus needs to continue avoiding the human before finally reaching its own goal at $t = 6.3s$. In contrast, the SLICE controller immediately recognizes that it can move close to its desired goal because it accurately predicts that the human will change its goal and move out of the way—thus staying both safe and live.

B. Ablation: When Does Modeling Influence Matter?

Here, ask the question: when does it matter that we inject an influence-aware model into the reach-avoid policy? We still assume the human agent plays according to the generated data distribution, but now try to understand the effect of choosing a *conditional* prediction model instead of a *marginal* one (i.e. one that does not condition on the robot's future plan). We compare both the performance of the predictors themselves and the downstream effect on the learned reach-avoid policies.

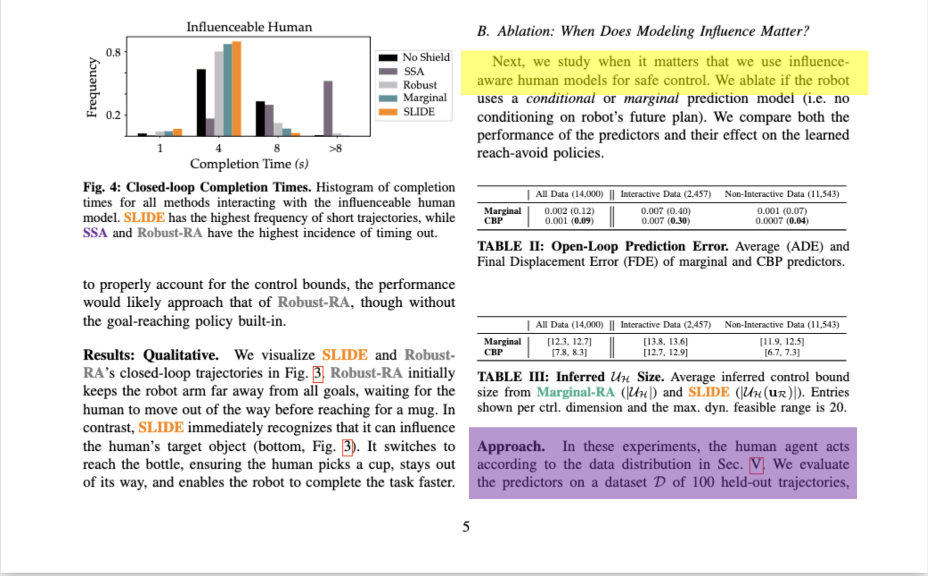
	All Data (14,000)	Interactive Data (2457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Average and Final Displacement Error of Marginal and CBP predictors reported as: ADE (FDE).

Open-Loop Results: Quantitative. The ADE is measured as the average L2 distance of the predicted end-effector position from the *first timestep* of the prediction horizon vs the true end-effector position. The FDE is measured as the L2 distance between the end-effector position predicted for the *final timestep* of the prediction horizon vs the true end effector position at the corresponding timestep. We test the predictors on a dataset of 100 held-out trajectories (generated in the same manner described in Sec. V), each 15 seconds long. This results in 14,000 data points[2], and as seen in Table II, the error rates look very similar, with the CBP model doing slightly better.

However, we know that highly interactive scenarios are long-tail events [9], so we decompose the datasets into data points where influence is happening $\mathcal{D}_I$ (2457 data points) and those where it is not $\mathcal{D}_{\neg I}$ (11,543 data points). This is done by tracking the timesteps where the human's goal changes—if it does, then this and directly adjacent timesteps are added to $\mathcal{D}_I$. We now uncover a stark difference in the FDE when

[2] We discount predictions in the last 1 second of each trajectory since this is length of the prediction horizon.



Influenceable Human
Frequency
0.8
0.2
1
4
8
>8
Completion Time (s)
No Shield
SSA
Robust
Marginal
SLIDE

Fig. 4: Closed-loop Completion Times. Histogram of completion times for all methods interacting with the influenceable human model. **SLIDE** has the highest frequency of short trajectories, while **SSA** and **Robust-RA** have the highest incidence of timing out.

to properly account for the control bounds, the performance would likely approach that of **Robust-RA**, though without the goal-reaching policy built-in.

Results: Qualitative. We visualize **SLIDE** and **Robust-RA**'s closed-loop trajectories in Fig. 3. **Robust-RA** initially keeps the robot arm far away from all goals, waiting for the human to move out of the way before reaching for a mug. In contrast, **SLIDE** immediately recognizes that it can influence the human's target object (bottom, Fig. 3). It switches to reach the bottle, ensuring the human picks a cup, stays out of its way, and enables the robot to complete the task faster.

B. Ablation: When Does Modeling Influence Matter?

Next, we study when it matters that we use influence-aware human models for safe control. We ablate if the robot uses a *conditional* or *marginal* prediction model (i.e. no conditioning on robot's future plan). We compare both the performance of the predictors and their effect on the learned reach-avoid policies.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	0.002 (0.12)	0.007 (0.40)	0.001 (0.07)
CBP	0.001 (**0.09**)	0.007 (**0.30**)	0.0007 (**0.04**)

TABLE II: Open-Loop Prediction Error. Average (ADE) and Final Displacement Error (FDE) of marginal and CBP predictors.

	All Data (14,000)	Interactive Data (2,457)	Non-Interactive Data (11,543)
Marginal	[12.3, 12.7]	[13.8, 13.6]	[11.9, 12.5]
CBP	[7.8, 8.3]	[12.7, 12.9]	[6.7, 7.3]

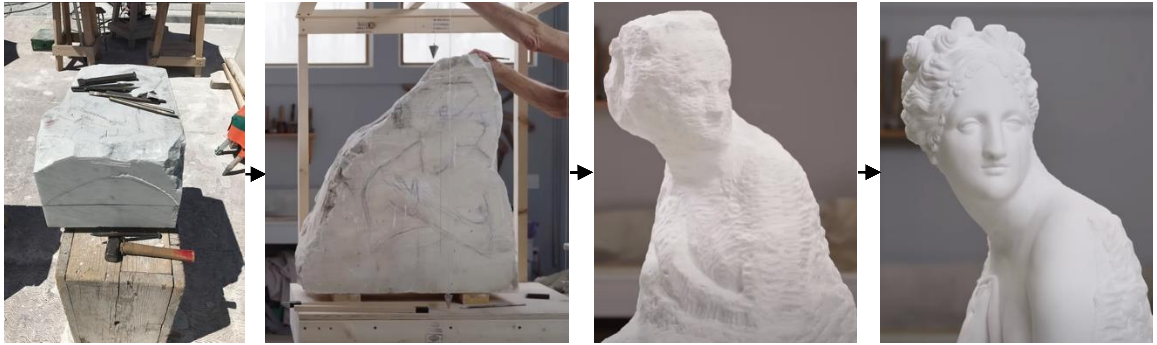
TABLE III: Inferred $\mathcal{U}_{\mathcal{H}}$ Size. Average inferred control bound size from **Marginal-RA** ($|\mathcal{U}_{\mathcal{H}}|$) and **SLIDE** ($|\mathcal{U}_{\mathcal{H}}(\mathbf{u}_{\mathcal{R}})|$). Entries shown per ctrl. dimension and the max. dyn. feasible range is 20.

Approach. In these experiments, the human agent acts according to the data distribution in Sec. V. We evaluate the predictors on a dataset $\mathcal{D}$ of 100 held-out trajectories,

Tighten the language – get to the point!

Clarify our experimental approach

Work Coarse to Fine-Grained



coarse ⟶ *fine*

Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

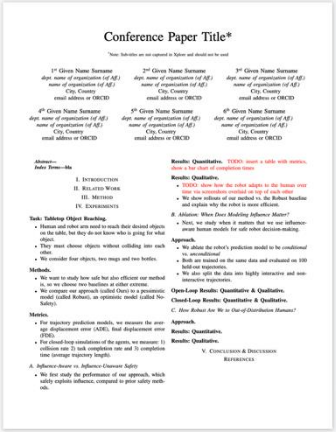
II. Related Work

III. Method

IV. Experiments

V. Conclusion & Discussion

References



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**
***Index Terms*—bla**

I. Introduction

II. Related Work

III. Method

IV. Experiments

Task: Tabletop Object Reaching.

- Human and robot arm need to reach their desired objects on the table, but they do not know who is going for what object.
- They must choose objects without colliding into each other.
- We consider four objects, two mugs and two bottles.

Methods.

- We want to study how safe but also efficient our method is, so we choose two baselines at either extreme.
- We compare our approach (called Ours) to a pessimistic model (called Robust), an optimistic model (called No-Safety).

Metrics.

- For trajectory prediction models, we measure the average displacement error (ADE), final displacement error (FDE).
- For closed-loop simulations of the agents, we measure: 1) collision rate 2) task completion rate and 3) completion time (average trajectory length).

A. Influence-Aware vs. Influence-Unaware Safety

- We first study the performance of our approach, which safely exploits influence, compared to prior safety methods.

Results: Quantitative. TODO: insert a table with metrics, show a bar chart of completion times

Results: Qualitative.

- TODO: show how the robot adapts to the human over time via screenshots overlaid on top of each other
- We show rollouts of our method vs. the Robust baseline and explain why the robot is more efficient.

B. Ablation: When Does Modeling Influence Matter?

- Next, we study when it matters that we use influence-aware human models for safe robot decision-making.

Approach.

- We ablate the robot's prediction model to be *conditional vs. unconditional*
- Both are trained on the same data and evaluated on 100 held-out trajectories.
- We also split the data into highly interactive and non-interactive trajectories.

Open-Loop Results: Quantitative & Qualitative.

Closed-Loop Results: Quantitative & Qualitative.

C. How Robust Are We to Out-of-Distribution Humans?

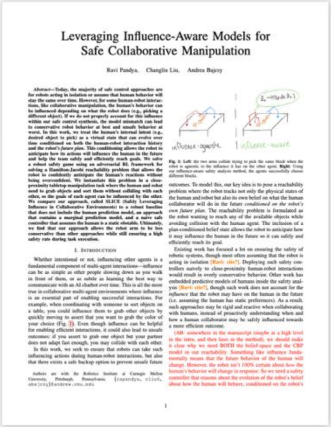
Approach.

Results: Quantitative.

Results: Qualitative.

V. Conclusion & Discussion

References



Leveraging Influence-Aware Models for Safe Collaborative Manipulation

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking a different object). If we do not properly account for this influence within our safe control synthesis, the model mismatch can lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired object to pick) as a virtual state that can evolve over time conditioned on both the human-robot interaction history and the robot's *future plan*. This conditioning allows the robot to anticipate how its actions will influence the human in the future and help the team safely and efficiently reach goals. We solve a robust safety game using an adversarial RL framework for solving a Hamilton-Jacobi reachability problem that allows the robot to confidently anticipate the human's reactions without being overconfident. We instantiate this problem in a close-proximity tabletop manipulation task where the human and robot need to grab objects and sort them without colliding with each other, so the goals of each agent can be influenced by the other. We compare our approach, called SLICE (Safely Leveraging Influence in Collaborative Environments) to a robust baseline that does not include the human prediction model, an approach that contains a marginal prediction model, and a naive safe controller that assumes the human is a static obstacle. Ultimately, we find that our approach allows the robot arm to be less conservative than other approaches while still ensuring a high safety rate during task execution.**

Fig. 1: Left: the two arms collide trying to pick the same block when the robot is agnostic to the influence it has on the other agent. **Right:** Using our influence-aware safety analysis method, the agents successfully choose different blocks.

I. Introduction

Whether intentional or not, influencing other agents is a fundamental component of multi-agent interactions—influence can be as simple as other people slowing down as you walk in front of them, or as subtle as learning the best way to communicate with an AI chatbot over time. This is all the more true in collaborative multi-agent environments where influence is an essential part of enabling successful interactions. For example, when coordinating with someone to sort objects on a table, you could influence them to grab other objects by quickly moving to assert that you want to grab the color of your choice (Fig. 1). Even though influence can be helpful for enabling efficient interactions, it could also lead to unsafe outcomes: if you assert to grab one object but your partner does not adapt fast enough, you may collide with each other.

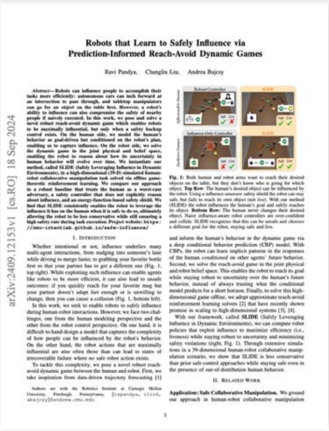
In this work, we seek to ensure that robots can take such influencing actions during human-robot interactions, but also that there exists a safe backup option to prevent unsafe future outcomes. To model this, our key idea is to pose a reachability problem where the robot tracks not only the physical states of the human and robot but also its own belief on what the human collaborator will do in the future *conditioned on the robot's own future plan*. The reachability problem is formulated as the robot wanting to reach any of the available objects while avoiding collision with the human agent. The inclusion of a plan-conditioned belief state allows the robot to anticipate how it may influence the human in the future so it can safely and efficiently reach its goal.

Existing work has focused a lot on ensuring the safety of robotic systems, though most often assuming that the robot is acting in isolation [Ravi: cite?]. Deploying such safety controllers naively to close-proximity human-robot interactions would result in overly conservative behavior. Other work has embedded predictive models of humans inside the safety analysis [Ravi: cite?], though such work does not account for the *influence* that the robot may have on the human in the future (i.e. assuming the human has static preferences). As a result, such approaches may be rigid and reactive when collaborating with humans, instead of proactively understanding when and how a human collaborator may be safely influenced towards a more efficient outcome.

(AB: somewhere in the manuscript (maybe at a high level in the intro, and then later in the method), we should make it clear why we need BOTH the belief-space and the CBF model in our reachability. Something like influence fundamentally means that the future behavior of the human will change. However, the robot isn't 100% certain about *how* the human's behavior will change in response. So we need a safety controller that reasons about the evolution of the robot's belief about how the human will behave, conditioned on the robot's

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1



arXiv:2409.12153v1 [cs.RO] 18 Sep 2024

Robots that Learn to Safely Influence via Prediction-Informed Reach-Avoid Dynamic Games

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*—Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row:** The human's desired object can be influenced by the robot. Using a influence-unaware safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row:** The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be unsafe and chooses a different goal for the robot, staying safe and live.

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from *nudging* into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where *no* safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' future behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

II. Related Work

Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. {rapandya, cliu6, abajcsy}@andrew.cmu.edu

1

Why does working “Coarse to Fine-Grained” help?

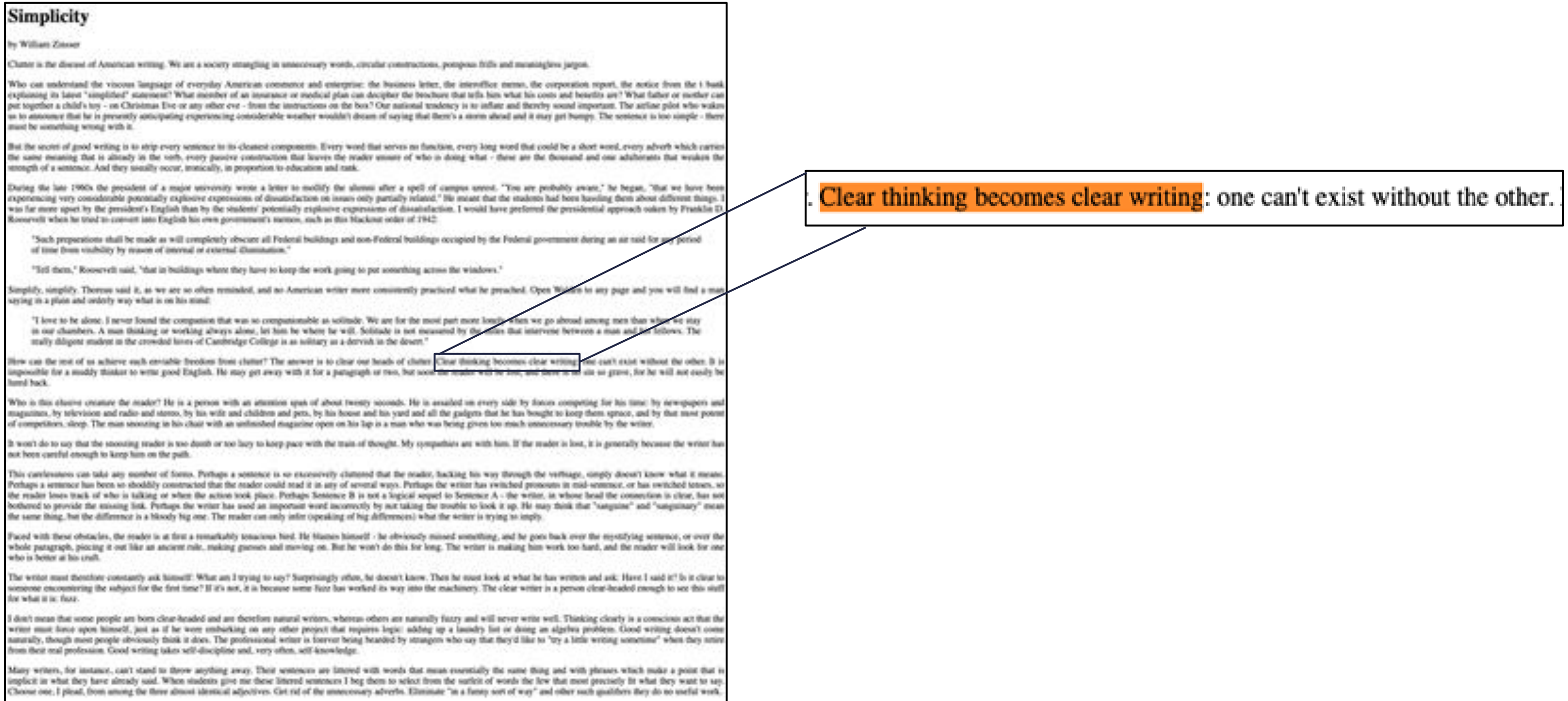
🤔

Clear thinking!

Theorem. Clear thinking == Clear writing.

Theorem. Clear thinking == Clear writing.

Proof. William Zinsser, W., 1980. Simplicity. In *"On Writing Well: An Informal Guide to Writing Nonfiction."*



Simplicity

by William Zinsser

Clutter is the disease of American writing. We are a society strangling in unnecessary words, circular constructions, pompous frills and meaningless jargon.

Who can understand the viscous language of everyday American commerce and enterprise: the business letter, the interoffice memo, the corporation report, the notice from the t bank explaining its latest "simplified" statement? What member of an insurance or medical plan can decipher the brochure that tells him what his costs and benefits are? What father or mother can put together a child's toy - on Christmas Eve or any other eve - from the instructions on the box? Our national tendency is to inflate and thereby sound important. The airline pilot who wakes us to announce that he is presently anticipating experiencing considerable weather wouldn't dream of saying that there's a storm ahead and it may get bumpy. The sentence is too simple - there must be something wrong with it.

But the secret of good writing is to strip every sentence to its cleanest components. Every word that serves no function, every long word that could be a short word, every adverb which carries the same meaning that is already in the verb, every passive construction that leaves the reader unsure of who is doing what - these are the thousand and one adulterants that weaken the strength of a sentence. And they usually occur, ironically, in proportion to education and rank.

During the late 1960s the president of a major university wrote a letter to mollify the alumni after a spell of campus unrest. "You are probably aware," he began, "that we have been experiencing very considerable potentially explosive expressions of dissatisfaction on issues only partially related." He meant that the students had been hassling them about different things. I was far more upset by the president's English than by the students' potentially explosive expressions of dissatisfaction. I would have preferred the presidential approach taken by Franklin D. Roosevelt when he tried to convert into English his own government's memos, such as this blackout order of 1942:

> "Such preparations shall be made as will completely obscure all Federal buildings and non-Federal buildings occupied by the Federal government during an air raid for any period of time from visibility by reason of internal or external illumination."

> "Tell them," Roosevelt said, "that in buildings where they have to keep the work going to put something across the windows."

Simplify, simplify. Thoreau said it, as we are so often reminded, and no American writer more consistently practiced what he preached. Open Walden to any page and you will find a man saying in a plain and orderly way what is on his mind:

> "I love to be alone. I never found the companion that was so companionable as solitude. We are for the most part more lonely when we go abroad among men than when we stay in our chambers. A man thinking or working always alone, let him be where he will. Solitude is not measured by the miles that intervene between a man and his fellows. The really diligent student in the crowded hives of Cambridge College is as solitary as a dervish in the desert."

How can the rest of us achieve such enviable freedom from clutter? The answer is to clear our heads of clutter. Clear thinking becomes clear writing: one can't exist without the other. It is impossible for a muddy thinker to write good English. He may get away with it for a paragraph or two, but soon the reader will be lost, and there is no sin so grave, for he will not easily be lured back.

Who is this elusive creature the reader? He is a person with an attention span of about twenty seconds. He is assailed on every side by forces competing for his time: by newspapers and magazines, by television and radio and stereo, by his wife and children and pets, by his house and his yard and all the gadgets that he has bought to keep them spruce, and by that most potent of competitors, sleep. The man snoozing in his chair with an unfinished magazine open on his lap is a man who was being given too much unnecessary trouble by the writer.

It won't do to say that the snoozing reader is too dumb or too lazy to keep pace with the train of thought. My sympathies are with him. If the reader is lost, it is generally because the writer has not been careful enough to keep him on the path.

This carelessness can take any number of forms. Perhaps a sentence is so excessively cluttered that the reader, hacking his way through the verbiage, simply doesn't know what it means. Perhaps a sentence has been so shoddily constructed that the reader could read it in any of several ways. Perhaps the writer has switched pronouns in mid-sentence, or has switched tenses, so the reader loses track of who is talking or when the action took place. Perhaps Sentence B is not a logical sequel to Sentence A - the writer, in whose head the connection is clear, has not bothered to provide the missing link. Perhaps the writer has used an important word incorrectly by not taking the trouble to look it up. He may think that "sanguine" and "sanguinary" mean the same thing, but the difference is a bloody big one. The reader can only infer (speaking of big differences) what the writer is trying to imply.

Faced with these obstacles, the reader is at first a remarkably tenacious bird. He blames himself - he obviously missed something, and he goes back over the mystifying sentence, or over the whole paragraph, piecing it out like an ancient rune, making guesses and moving on. But he won't do this for long. The writer is making him work too hard, and the reader will look for one who is better at his craft.

The writer must therefore constantly ask himself: What am I trying to say? Surprisingly often, he doesn't know. Then he must look at what he has written and ask: Have I said it? Is it clear to someone encountering the subject for the first time? If it's not, it is because some fuzz has worked its way into the machinery. The clear writer is a person clear-headed enough to see this stuff for what it is: fuzz.

I don't mean that some people are born clear-headed and are therefore natural writers, whereas others are naturally fuzzy and will never write well. Thinking clearly is a conscious act that the writer must force upon himself, just as if he were embarking on any other project that requires logic: adding up a laundry list or doing an algebra problem. Good writing doesn't come naturally, though most people obviously think it does. The professional writer is forever being bearded by strangers who say that they'd like to "try a little writing sometime" when they retire from their real profession. Good writing takes self-discipline and, very often, self-knowledge.

Many writers, for instance, can't stand to throw anything away. Their sentences are littered with words that mean essentially the same thing and with phrases which make a point that is implicit in what they have already said. When students give me these littered sentences I beg them to select from the surfeit of words the few that most precisely fit what they want to say. Choose one, I plead, from among the three almost identical adjectives. Get rid of the unnecessary adverbs. Eliminate "in a funny sort of way" and other such qualifiers they do no useful work.

. Clear thinking becomes clear writing: one can't exist without the other.

Theorem. Clear thinking == Clear writing.

Proof. William Zinsser, W., 1980. Simplicity. In *"On Writing Well: An Informal Guide to Writing Nonfiction."*

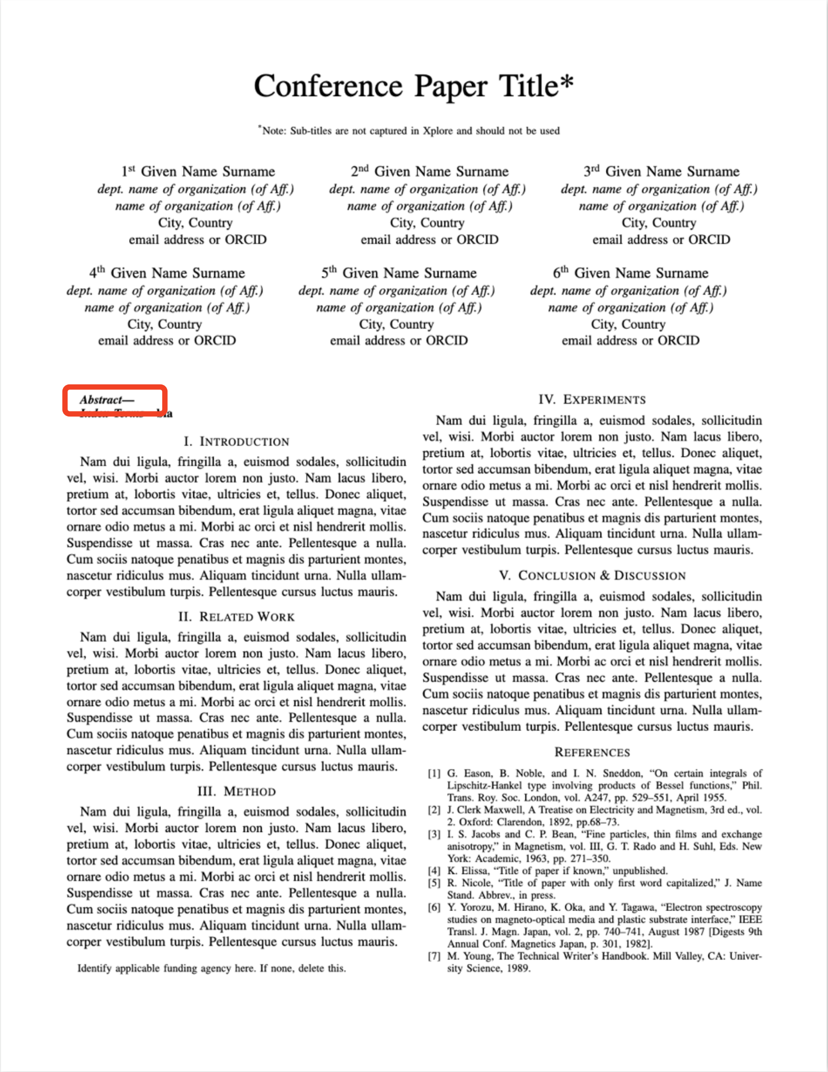
Now: Reading ☺

Clear thinking is particularly helpful when it comes to the *abstract* and *introductions*

These sections require the most "distillation" of your key ideas

Reviewers will often decide at a "gut level" if a paper should be accepted or rejected based on the introduction!

Abstract



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—ia**

I. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

II. Related Work

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

III. Method

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

IV. Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

V. Conclusion & Discussion

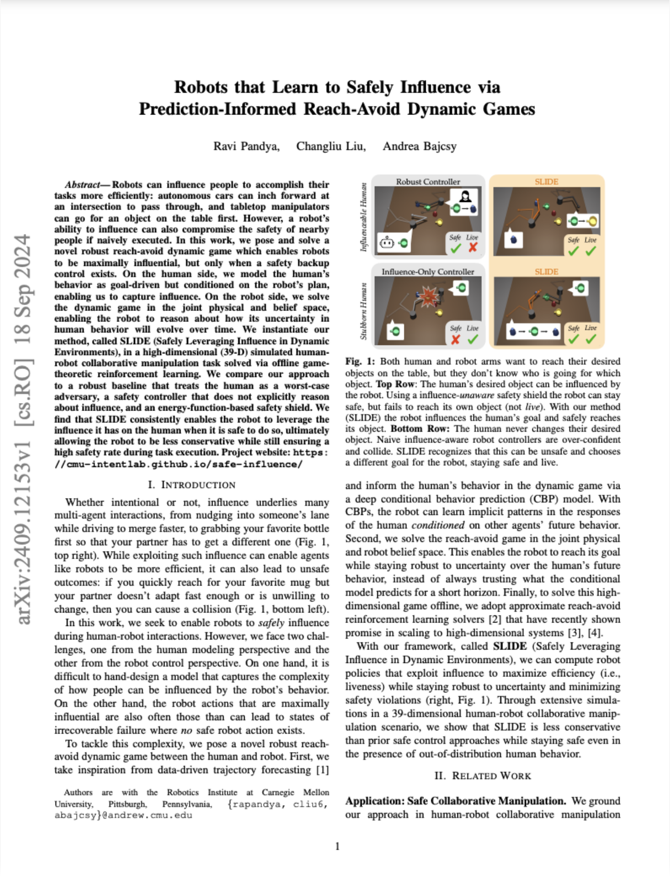
Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, “Title of paper if known,” unpublished.
[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Identify applicable funding agency here. If none, delete this.

Example (ICRA 2025 Submission)



Robots that Learn to Safely Influence via Prediction-Informed Reach-Avoid Dynamic Games

Ravi Pandya, Changliu Liu, Andrea Bajcsy

***Abstract*— Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution. Project website: `https://cmu-intentlab.github.io/safe-influence/`**

Fig. 1: Both human and robot arms want to reach their desired objects on the table, but they don't know who is going for which object. **Top Row**: The human's desired object can be influenced by the robot. Using a influence-*unaware* safety shield the robot can stay safe, but fails to reach its own object (not *live*). With our method (SLIDE) the robot influences the human's goal and safely reaches its object. **Bottom Row:** The human never changes their desired object. Naive influence-aware robot controllers are over-confident and collide. SLIDE recognizes that this can be unsafe and chooses a different goal for the robot, staying safe and live.

I. Introduction

Whether intentional or not, influence underlies many multi-agent interactions, from nudging into someone's lane while driving to merge faster, to grabbing your favorite bottle first so that your partner has to get a different one (Fig. 1, top right). While exploiting such influence can enable agents like robots to be more efficient, it can also lead to unsafe outcomes: if you quickly reach for your favorite mug but your partner doesn't adapt fast enough or is unwilling to change, then you can cause a collision (Fig. 1, bottom left).

In this work, we seek to enable robots to *safely* influence during human-robot interactions. However, we face two challenges, one from the human modeling perspective and the other from the robot control perspective. On one hand, it is difficult to hand-design a model that captures the complexity of how people can be influenced by the robot's behavior. On the other hand, the robot actions that are maximally influential are also often those than can lead to states of irrecoverable failure where *no* safe robot action exists.

To tackle this complexity, we pose a novel robust reach-avoid dynamic game between the human and robot. First, we take inspiration from data-driven trajectory forecasting [1] and inform the human's behavior in the dynamic game via a deep conditional behavior prediction (CBP) model. With CBPs, the robot can learn implicit patterns in the responses of the human *conditioned* on other agents' future behavior. Second, we solve the reach-avoid game in the joint physical and robot belief space. This enables the robot to reach its goal while staying robust to uncertainty over the human's future behavior, instead of always trusting what the conditional model predicts for a short horizon. Finally, to solve this high-dimensional game offline, we adopt approximate reach-avoid reinforcement learning solvers [2] that have recently shown promise in scaling to high-dimensional systems [3], [4].

With our framework, called **SLIDE** (Safely Leveraging Influence in Dynamic Environments), we can compute robot policies that exploit influence to maximize efficiency (i.e., liveness) while staying robust to uncertainty and minimizing safety violations (right, Fig. 1). Through extensive simulations in a 39-dimensional human-robot collaborative manipulation scenario, we show that SLIDE is less conservative than prior safe control approaches while staying safe even in the presence of out-of-distribution human behavior.

II. Related Work

Application: Safe Collaborative Manipulation. We ground our approach in human-robot collaborative manipulation

Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania, `{rapandya, cliu6, abajcsy}@andrew.cmu.edu`

Initial Abstract

Tells us about an assumption

Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking an alternative block). If we do not properly account for this influence within our safe control synthesis, the model mismatch will lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired block to pick) as a virtual state which evolves over time as a function of the past human-robot interaction history. By augmenting the robot's state space with the time-varying human intent, we can solve a robust safety game that enables the robot to anticipate how present actions will influence the human to change their goal and, in turn, bring the human-robot team closer to an unsafe state. We instantiate this idea in a close-proximity tabletop manipulation task where the human and robot have to efficiently grab blocks and sort them without colliding into each other, but who grabs which block can be influenced by the other agent. By using a learned model of how the human collaborator's goal can change over time within our safety analysis framework, the robotic arm is less conservative while still ensuring a high safety rate during task execution.

Question: What's **good**?

Tells us what goes wrong under assumption

Describes takeaways

Initial Abstract

Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking an alternative block). If we do not properly account for this influence within our safe control synthesis, the model mismatch will lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired block to pick) as a virtual state which evolves over time as a function of the past human-robot interaction history. By augmenting the robot's state space with the time-varying human intent, we can solve a robust safety game that enables the robot to anticipate how present actions will influence the human to change their goal and, in turn, bring the human-robot team closer to an unsafe state. We instantiate this idea in a close-proximity tabletop manipulation task where the human and robot have to efficiently grab blocks and sort them without colliding into each other, but who grabs which block can be influenced by the other agent. By using a learned model of how the human collaborator's goal can change over time within our safety analysis framework, the robotic arm is less conservative while still ensuring a high safety rate during task execution.

Question: What's **wrong**?

Contribution isn't the clearest

Have to read ½ way to learn what we did!

How did we solve this robust safety game?

What is this improvement relative to?

Initial Abstract

Today, the majority of safe control approaches are for robots acting in isolation or assume that human behavior will stay the same over time. However, for some human-robot interactions, like collaborative manipulation, the human's behavior can be influenced depending on what the robot does (e.g., picking an alternative block). If we do not properly account for this influence within our safe control synthesis, the model mismatch will lead to conservative robot behavior at best and unsafe behavior at worst. In this work, we treat the human's internal intent (e.g., desired block to pick) as a virtual state which evolves over time as a function of the past human-robot interaction history. By augmenting the robot's state space with the time-varying human intent, we can solve a robust safety game that enables the robot to anticipate how present actions will influence the human to change their goal and, in turn, bring the human-robot team closer to an unsafe state. We instantiate this idea in a close-proximity tabletop manipulation task where the human and robot have to efficiently grab blocks and sort them without colliding into each other, but who grabs which block can be influenced by the other agent. By using a learned model of how the human collaborator's goal can change over time within our safety analysis framework, the robotic arm is less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Key contribution stated up-front

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Key contribution stated up-front

Two aspects of contributions highlighted

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Key contribution stated up-front

Two aspects of contributions highlighted

How we accomplish / test our idea

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Key contribution stated up-front

Two aspects of contributions highlighted

How we accomplish / test our idea

What we compare to

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Final Abstract

Tells us the problem setting & gap

Key contribution stated up-front

Two aspects of contributions highlighted

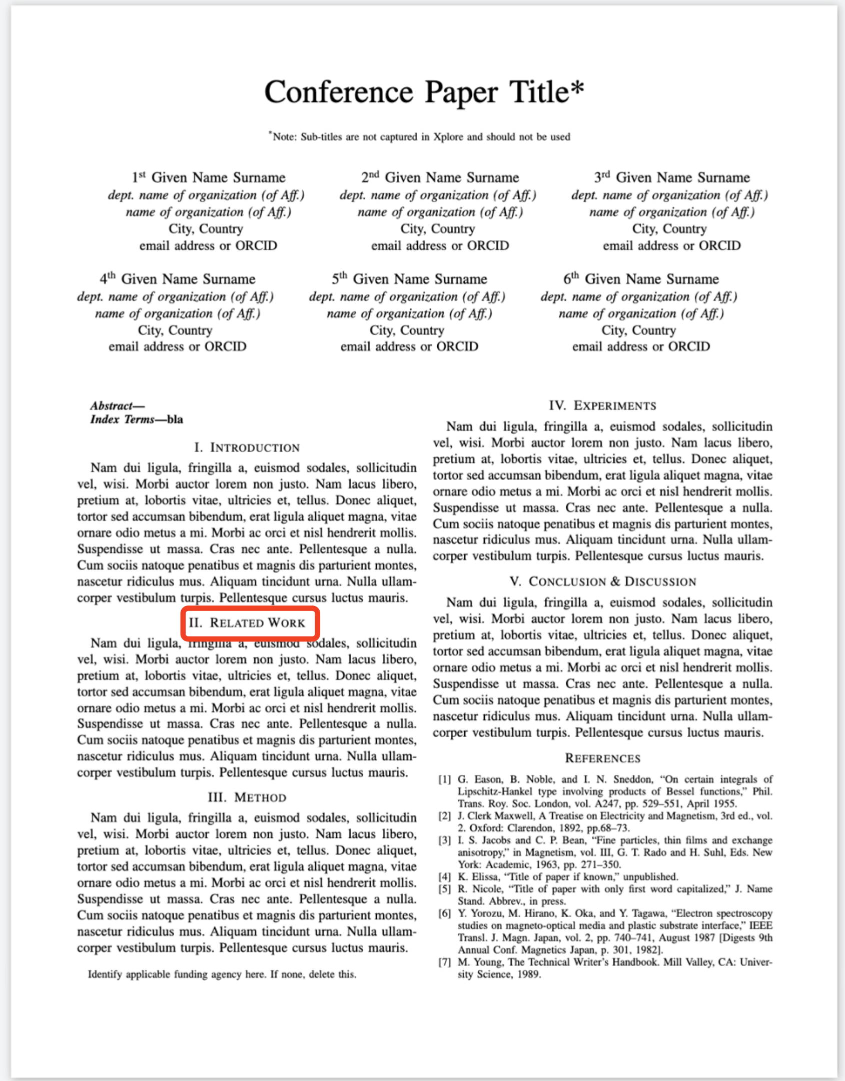
How we accomplish / test our idea

What we compare to

Takeaways from empirical results

Robots can influence people to accomplish their tasks more efficiently: autonomous cars can inch forward at an intersection to pass through, and tabletop manipulators can go for an object on the table first. However, a robot's ability to influence can also compromise the safety of nearby people if naively executed. In this work, we pose and solve a novel robust reach-avoid dynamic game which enables robots to be maximally influential, but only when a safety backup control exists. On the human side, we model the human's behavior as goal-driven but conditioned on the robot's plan, enabling us to capture influence. On the robot side, we solve the dynamic game in the joint physical and belief space, enabling the robot to reason about how its uncertainty in human behavior will evolve over time. We instantiate our method, called SLIDE (Safely Leveraging Influence in Dynamic Environments), in a high-dimensional (39-D) simulated human-robot collaborative manipulation task solved via offline game-theoretic reinforcement learning. We compare our approach to a robust baseline that treats the human as a worst-case adversary, a safety controller that does not explicitly reason about influence, and an energy-function-based safety shield. We find that SLIDE consistently enables the robot to leverage the influence it has on the human when it is safe to do so, ultimately allowing the robot to be less conservative while still ensuring a high safety rate during task execution.

Related Work



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—bla**

I. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

II. Related Work

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

III. Method

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

IV. Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

V. Conclusion & Discussion

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, “Title of paper if known,” unpublished.
[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

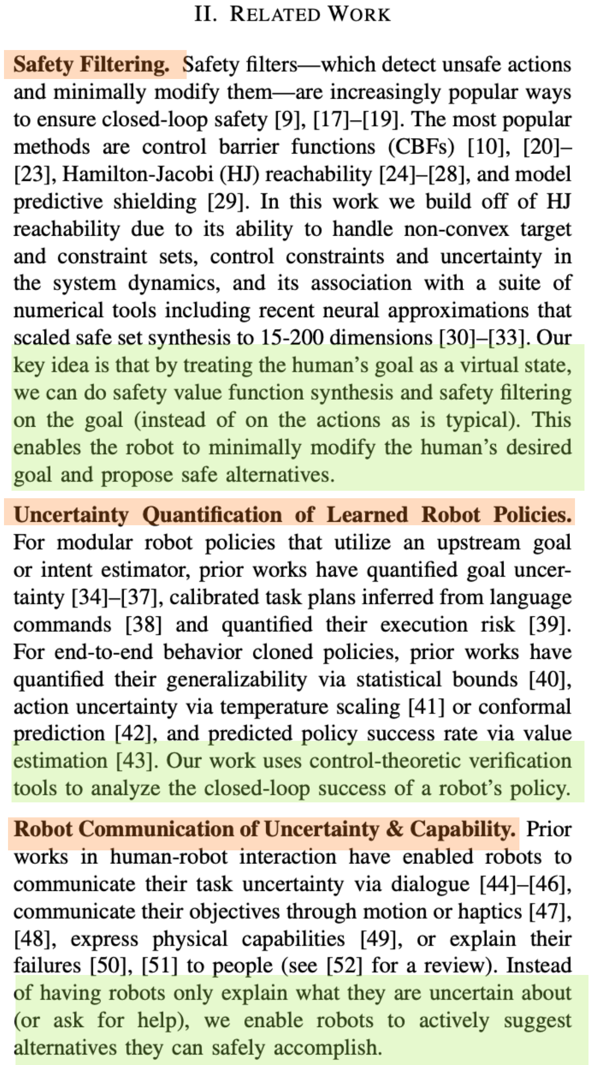
Identify applicable funding agency here. If none, delete this.

Related works are *not:*

- ⊗ lists
- ⊗ every single paper you came across during research
- ⊗ disconnected from the other sections

Related works *are*:

- ☑ descriptions of where we “are” in a field
- ☑ opportunities to highlight open gaps your work solves
- ☑ relevant to the key “dimensions” of your work



II. Related Work

Safety Filtering. Safety filters—which detect unsafe actions and minimally modify them—are increasingly popular ways to ensure closed-loop safety [9], [17]–[19]. The most popular methods are control barrier functions (CBFs) [10], [20]–[23], Hamilton-Jacobi (HJ) reachability [24]–[28], and model predictive shielding [29]. In this work we build off of HJ reachability due to its ability to handle non-convex target and constraint sets, control constraints and uncertainty in the system dynamics, and its association with a suite of numerical tools including recent neural approximations that scaled safe set synthesis to 15-200 dimensions [30]–[33]. Our key idea is that by treating the human's goal as a virtual state, we can do safety value function synthesis and safety filtering on the goal (instead of on the actions as is typical). This enables the robot to minimally modify the human's desired goal and propose safe alternatives.

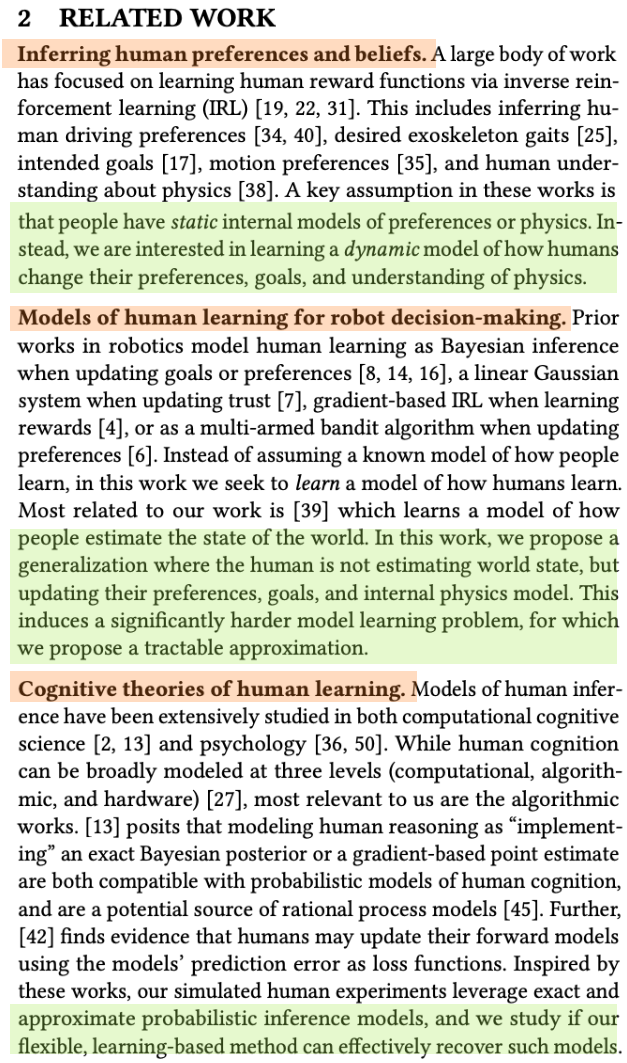
Uncertainty Quantification of Learned Robot Policies. For modular robot policies that utilize an upstream goal or intent estimator, prior works have quantified goal uncertainty [34]–[37], calibrated task plans inferred from language commands [38] and quantified their execution risk [39]. For end-to-end behavior cloned policies, prior works have quantified their generalizability via statistical bounds [40], action uncertainty via temperature scaling [41] or conformal prediction [42], and predicted policy success rate via value estimation [43]. Our work uses control-theoretic verification tools to analyze the closed-loop success of a robot's policy.

Robot Communication of Uncertainty & Capability. Prior works in human-robot interaction have enabled robots to communicate their task uncertainty via dialogue [44]–[46], communicate their objectives through motion or haptics [47], [48], express physical capabilities [49], or explain their failures [50], [51] to people (see [52] for a review). Instead of having robots only explain what they are uncertain about (or ask for help), we enable robots to actively suggest alternatives they can safely accomplish.

Sections highlight the key "dimensions"

End of each section highlights what is new about our approach from this "dimension"

Content describes ***foundational*** *work as well as* ***recent*** *work that is most relevant*



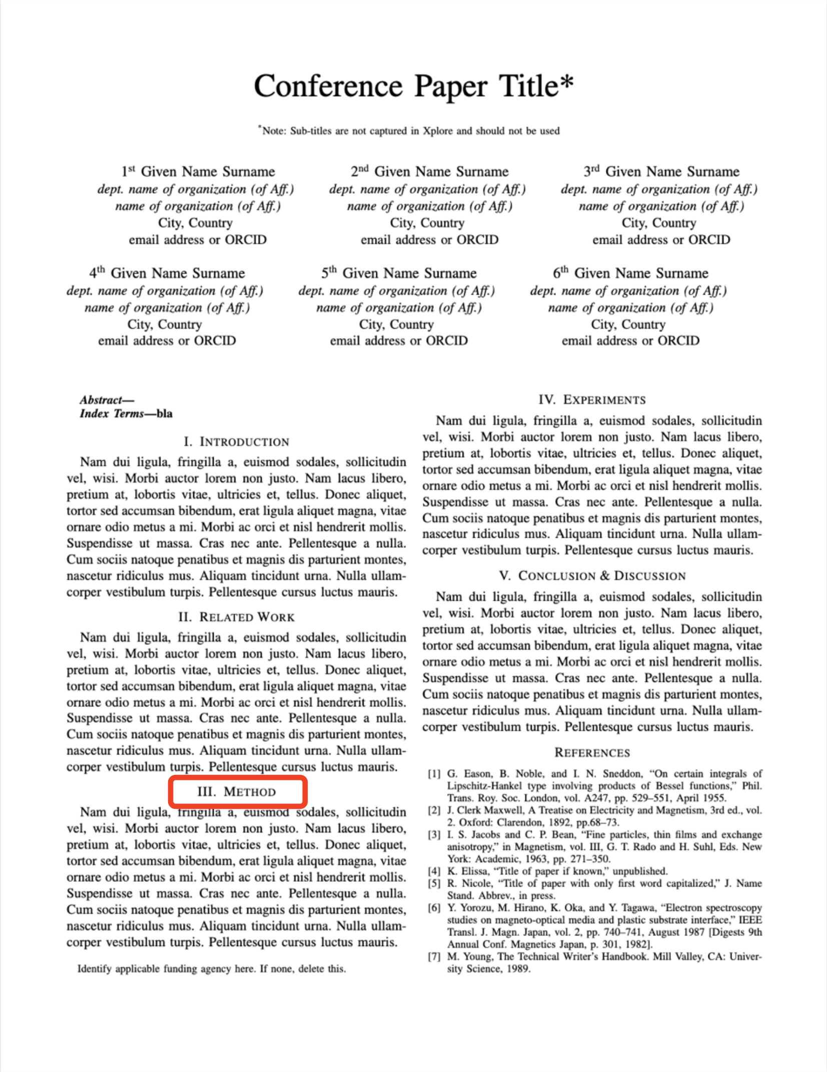
2 RELATED WORK

Inferring human preferences and beliefs. A large body of work has focused on learning human reward functions via inverse reinforcement learning (IRL) [19, 22, 31]. This includes inferring human driving preferences [34, 40], desired exoskeleton gaits [25], intended goals [17], motion preferences [35], and human understanding about physics [38]. A key assumption in these works is that people have *static* internal models of preferences or physics. Instead, we are interested in learning a *dynamic* model of how humans change their preferences, goals, and understanding of physics.

Models of human learning for robot decision-making. Prior works in robotics model human learning as Bayesian inference when updating goals or preferences [8, 14, 16], a linear Gaussian system when updating trust [7], gradient-based IRL when learning rewards [4], or as a multi-armed bandit algorithm when updating preferences [6]. Instead of assuming a known model of how people learn, in this work we seek to *learn* a model of how humans learn. Most related to our work is [39] which learns a model of how people estimate the state of the world. In this work, we propose a generalization where the human is not estimating world state, but updating their preferences, goals, and internal physics model. This induces a significantly harder model learning problem, for which we propose a tractable approximation.

Cognitive theories of human learning. Models of human inference have been extensively studied in both computational cognitive science [2, 13] and psychology [36, 50]. While human cognition can be broadly modeled at three levels (computational, algorithmic, and hardware) [27], most relevant to us are the algorithmic works. [13] posits that modeling human reasoning as "implementing" an exact Bayesian posterior or a gradient-based point estimate are both compatible with probabilistic models of human cognition, and are a potential source of rational process models [45]. Further, [42] finds evidence that humans may update their forward models using the models' prediction error as loss functions. Inspired by these works, our simulated human experiments leverage exact and approximate probabilistic inference models, and we study if our flexible, learning-based method can effectively recover such models.

Method



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—bla**

I. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

II. Related Work

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

III. Method

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

IV. Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

V. Conclusion & Discussion

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

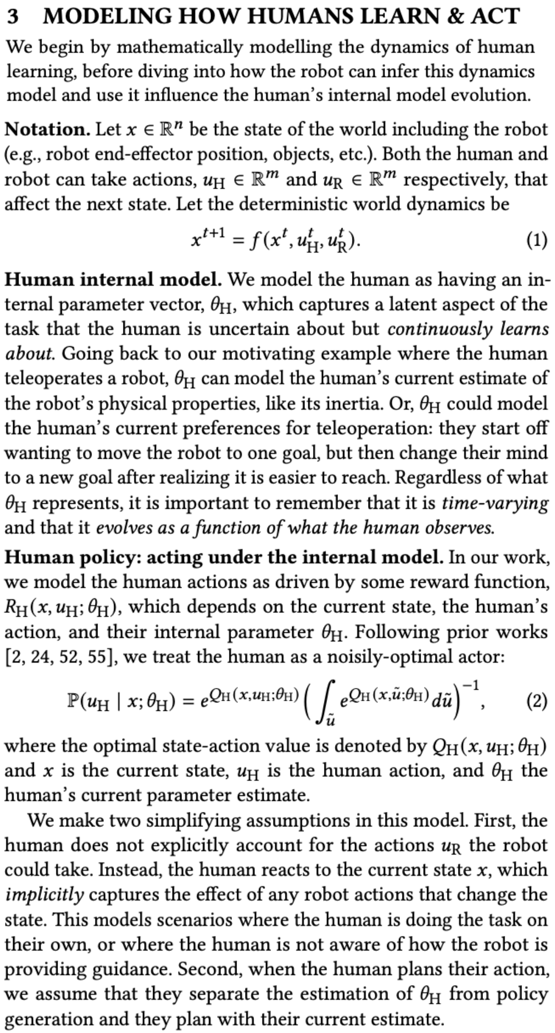
References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, “Title of paper if known,” unpublished.
[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Identify applicable funding agency here. If none, delete this.

Example 1

(algorithmic contribution)



3 MODELING HOW HUMANS LEARN & ACT

We begin by mathematically modelling the dynamics of human learning, before diving into how the robot can infer this dynamics model and use it influence the human's internal model evolution.

Notation. Let $x \in \mathbb{R}^n$ be the state of the world including the robot (e.g., robot end-effector position, objects, etc.). Both the human and robot can take actions, $u_\mathrm{H} \in \mathbb{R}^m$ and $u_\mathrm{R} \in \mathbb{R}^m$ respectively, that affect the next state. Let the deterministic world dynamics be

$$x^{t+1} = f(x^t, u_\mathrm{H}^t, u_\mathrm{R}^t). \quad (1)$$

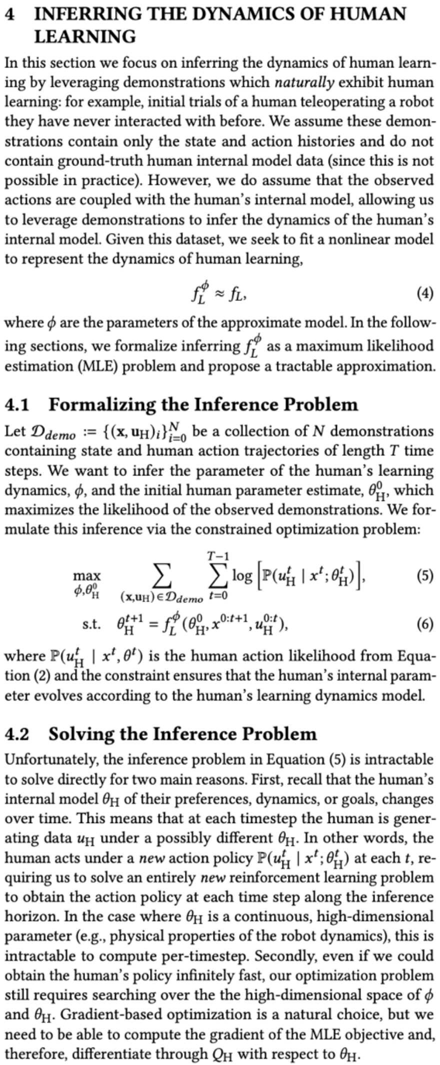
Human internal model. We model the human as having an internal parameter vector, θ_H, which captures a latent aspect of the task that the human is uncertain about but *continuously learns about.* Going back to our motivating example where the human teleoperates a robot, θ_H can model the human's current estimate of the robot's physical properties, like its inertia. Or, θ_H could model the human's current preferences for teleoperation: they start off wanting to move the robot to one goal, but then change their mind to a new goal after realizing it is easier to reach. Regardless of what θ_H represents, it is important to remember that it is *time-varying* and that it *evolves as a function of what the human observes.*

Human policy: acting under the internal model. In our work, we model the human actions as driven by some reward function, $R_\mathrm{H}(x, u_\mathrm{H}; \theta_\mathrm{H})$, which depends on the current state, the human's action, and their internal parameter θ_H. Following prior works [2, 24, 52, 55], we treat the human as a noisily-optimal actor:

$$\mathbb{P}(u_\mathrm{H} \mid x; \theta_\mathrm{H}) = e^{Q_\mathrm{H}(x,u_\mathrm{H};\theta_\mathrm{H})} \left(\int_{\tilde{u}} e^{Q_\mathrm{H}(x,\tilde{u};\theta_\mathrm{H})} d\tilde{u} \right)^{-1}, \quad (2)$$

where the optimal state-action value is denoted by $Q_\mathrm{H}(x, u_\mathrm{H}; \theta_\mathrm{H})$ and x is the current state, u_H is the human action, and θ_H the human's current parameter estimate.

We make two simplifying assumptions in this model. First, the human does not explicitly account for the actions u_R the robot could take. Instead, the human reacts to the current state x, which *implicitly* captures the effect of any robot actions that change the state. This models scenarios where the human is doing the task on their own, or where the human is not aware of how the robot is providing guidance. Second, when the human plans their action, we assume that they separate the estimation of θ_H from policy generation and they plan with their current estimate.



4 INFERRING THE DYNAMICS OF HUMAN LEARNING

In this section we focus on inferring the dynamics of human learning by leveraging demonstrations which *naturally* exhibit human learning: for example, initial trials of a human teleoperating a robot they have never interacted with before. We assume these demonstrations contain only the state and action histories and do not contain ground-truth human internal model data (since this is not possible in practice). However, we do assume that the observed actions are coupled with the human's internal model, allowing us to leverage demonstrations to infer the dynamics of the human's internal model. Given this dataset, we seek to fit a nonlinear model to represent the dynamics of human learning,

$$f_L^\phi \approx f_L, \quad (4)$$

where ϕ are the parameters of the approximate model. In the following sections, we formalize inferring f_L^ϕ as a maximum likelihood estimation (MLE) problem and propose a tractable approximation.

4.1 Formalizing the Inference Problem

Let $\mathcal{D}_{demo} := \{(\mathbf{x}, \mathbf{u}_\mathrm{H})_i\}_{i=0}^N$ be a collection of N demonstrations containing state and human action trajectories of length T time steps. We want to infer the parameter of the human's learning dynamics, ϕ, and the initial human parameter estimate, θ_H^0, which maximizes the likelihood of the observed demonstrations. We formulate this inference via the constrained optimization problem:

$$\max_{\phi,\theta_\mathrm{H}^0} \sum_{(\mathbf{x},\mathbf{u}_\mathrm{H}) \in \mathcal{D}_{demo}} \sum_{t=0}^{T-1} \log \left[\mathbb{P}(u_\mathrm{H}^t \mid x^t; \theta_\mathrm{H}^t) \right], \quad (5)$$

$$\text{s.t.} \quad \theta_\mathrm{H}^{t+1} = f_L^\phi(\theta_\mathrm{H}^0, x^{0:t+1}, u_\mathrm{H}^{0:t}), \quad (6)$$

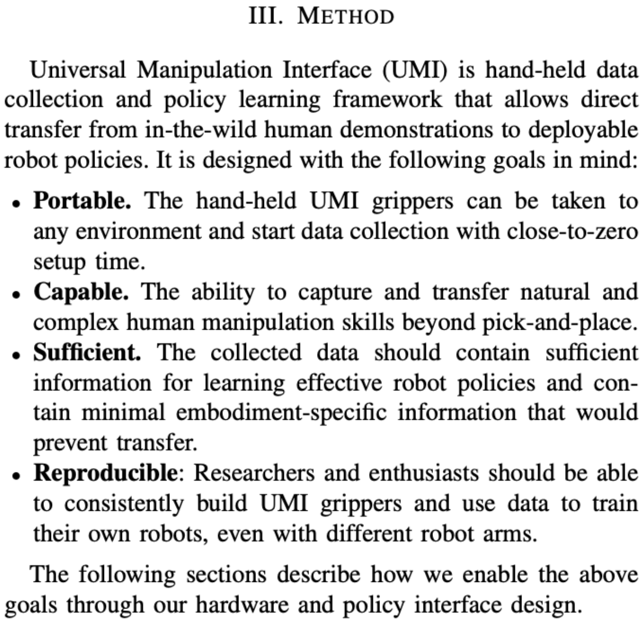
where $\mathbb{P}(u_\mathrm{H}^t \mid x^t, \theta^t)$ is the human action likelihood from Equation (2) and the constraint ensures that the human's internal parameter evolves according to the human's learning dynamics model.

4.2 Solving the Inference Problem

Unfortunately, the inference problem in Equation (5) is intractable to solve directly for two main reasons. First, recall that the human's internal model θ_H of their preferences, dynamics, or goals, changes over time. This means that at each timestep the human is generating data u_H under a possibly different θ_H. In other words, the human acts under a *new* action policy $\mathbb{P}(u_\mathrm{H}^t \mid x^t; \theta_\mathrm{H}^t)$ at each t, requiring us to solve an entirely *new* reinforcement learning problem to obtain the action policy at each time step along the inference horizon. In the case where θ_H is a continuous, high-dimensional parameter (e.g., physical properties of the robot dynamics), this is intractable to compute per-timestep. Secondly, even if we could obtain the human's policy infinitely fast, our optimization problem still requires searching over the the high-dimensional space of ϕ and θ_H. Gradient-based optimization is a natural choice, but we need to be able to compute the gradient of the MLE objective and, therefore, differentiate through Q_H with respect to θ_H.

Example 2

(system contribution)

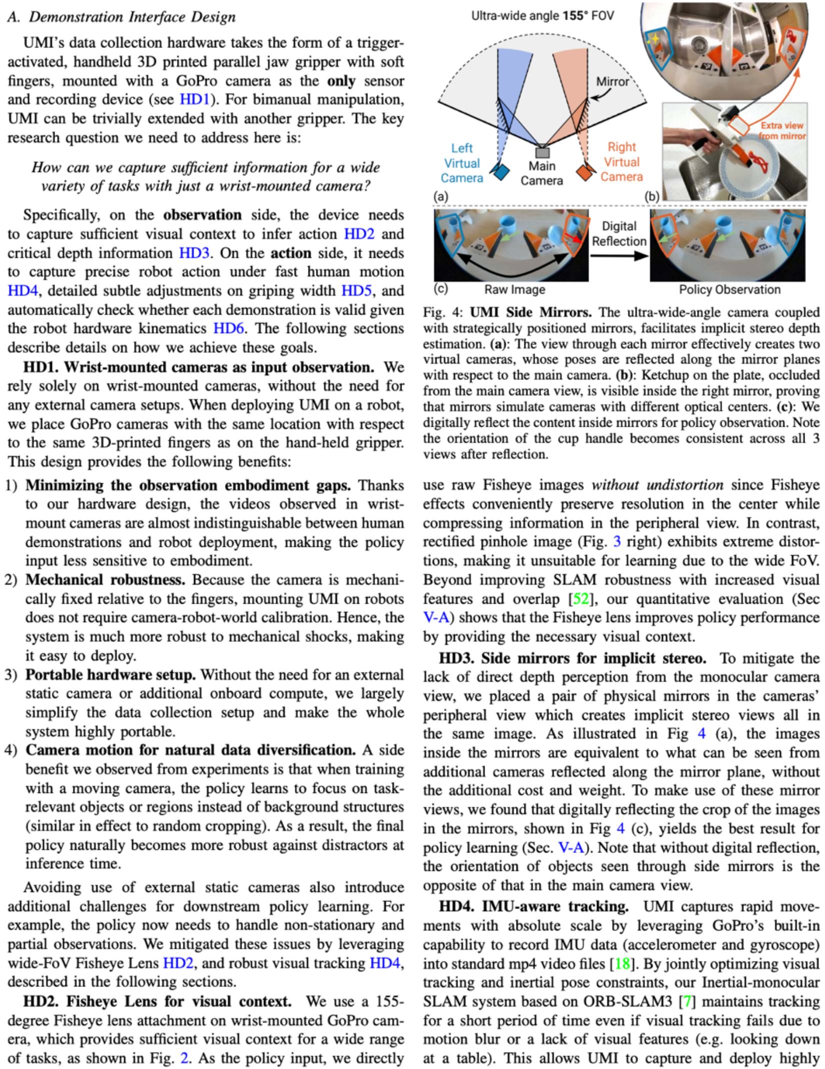


III. Method

Universal Manipulation Interface (UMI) is hand-held data collection and policy learning framework that allows direct transfer from in-the-wild human demonstrations to deployable robot policies. It is designed with the following goals in mind:

- **Portable.** The hand-held UMI grippers can be taken to any environment and start data collection with close-to-zero setup time.
- **Capable.** The ability to capture and transfer natural and complex human manipulation skills beyond pick-and-place.
- **Sufficient.** The collected data should contain sufficient information for learning effective robot policies and contain minimal embodiment-specific information that would prevent transfer.
- **Reproducible**: Researchers and enthusiasts should be able to consistently build UMI grippers and use data to train their own robots, even with different robot arms.

The following sections describe how we enable the above goals through our hardware and policy interface design.



A. Demonstration Interface Design

UMI's data collection hardware takes the form of a trigger-activated, handheld 3D printed parallel jaw gripper with soft fingers, mounted with a GoPro camera as the **only** sensor and recording device (see HD1). For bimanual manipulation, UMI can be trivially extended with another gripper. The key research question we need to address here is:

How can we capture sufficient information for a wide variety of tasks with just a wrist-mounted camera?

Specifically, on the **observation** side, the device needs to capture sufficient visual context to infer action HD2 and critical depth information HD3. On the **action** side, it needs to capture precise robot action under fast human motion HD4, detailed subtle adjustments on griping width HD5, and automatically check whether each demonstration is valid given the robot hardware kinematics HD6. The following sections describe details on how we achieve these goals.

HD1. Wrist-mounted cameras as input observation. We rely solely on wrist-mounted cameras, without the need for any external camera setups. When deploying UMI on a robot, we place GoPro cameras with the same location with respect to the same 3D-printed fingers as on the hand-held gripper. This design provides the following benefits:

1) **Minimizing the observation embodiment gaps.** Thanks to our hardware design, the videos observed in wrist-mount cameras are almost indistinguishable between human demonstrations and robot deployment, making the policy input less sensitive to embodiment.
2) **Mechanical robustness.** Because the camera is mechanically fixed relative to the fingers, mounting UMI on robots does not require camera-robot-world calibration. Hence, the system is much more robust to mechanical shocks, making it easy to deploy.
3) **Portable hardware setup.** Without the need for an external static camera or additional onboard compute, we largely simplify the data collection setup and make the whole system highly portable.
4) **Camera motion for natural data diversification.** A side benefit we observed from experiments is that when training with a moving camera, the policy learns to focus on task-relevant objects or regions instead of background structures (similar in effect to random cropping). As a result, the final policy naturally becomes more robust against distractors at inference time.

Avoiding use of external static cameras also introduce additional challenges for downstream policy learning. For example, the policy now needs to handle non-stationary and partial observations. We mitigated these issues by leveraging wide-FoV Fisheye Lens HD2, and robust visual tracking HD4, described in the following sections.

HD2. Fisheye Lens for visual context. We use a 155-degree Fisheye lens attachment on wrist-mounted GoPro camera, which provides sufficient visual context for a wide range of tasks, as shown in Fig. 2. As the policy input, we directly use raw Fisheye images *without undistortion* since Fisheye effects conveniently preserve resolution in the center while compressing information in the peripheral view. In contrast, rectified pinhole image (Fig. 3 right) exhibits extreme distortions, making it unsuitable for learning due to the wide FoV. Beyond improving SLAM robustness with increased visual features and overlap [52], our quantitative evaluation (Sec V-A) shows that the Fisheye lens improves policy performance by providing the necessary visual context.

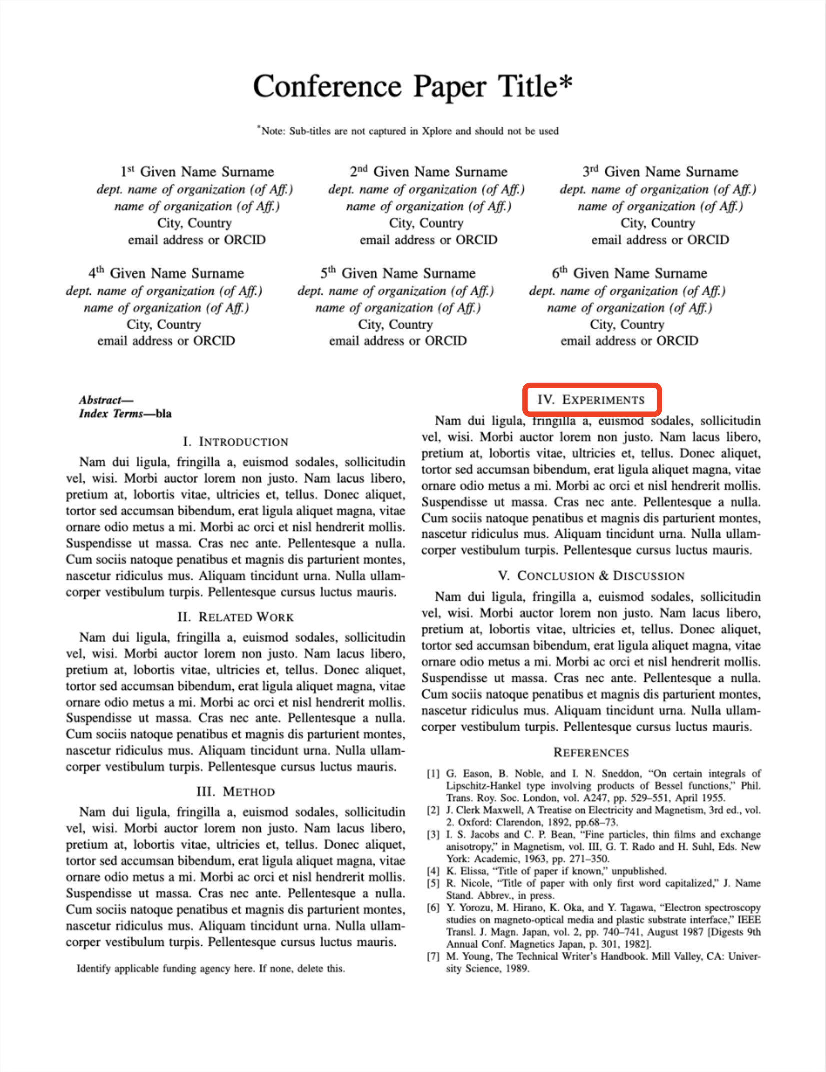
Fig. 4: **UMI Side Mirrors.** The ultra-wide-angle camera coupled with strategically positioned mirrors, facilitates implicit stereo depth estimation. **(a)**: The view through each mirror effectively creates two virtual cameras, whose poses are reflected along the mirror planes with respect to the main camera. **(b)**: Ketchup on the plate, occluded from the main camera view, is visible inside the right mirror, proving that mirrors simulate cameras with different optical centers. **(c)**: We digitally reflect the content inside mirrors for policy observation. Note the orientation of the cup handle becomes consistent across all 3 views after reflection.

HD3. Side mirrors for implicit stereo. To mitigate the lack of direct depth perception from the monocular camera view, we placed a pair of physical mirrors in the cameras' peripheral view which creates implicit stereo views all in the same image. As illustrated in Fig 4 (a), the images inside the mirrors are equivalent to what can be seen from additional cameras reflected along the mirror plane, without the additional cost and weight. To make use of these mirror views, we found that digitally reflecting the crop of the images in the mirrors, shown in Fig 4 (c), yields the best result for policy learning (Sec. V-A). Note that without digital reflection, the orientation of objects seen through side mirrors is the opposite of that in the main camera view.

HD4. IMU-aware tracking. UMI captures rapid movements with absolute scale by leveraging GoPro's built-in capability to record IMU data (accelerometer and gyroscope) into standard mp4 video files [18]. By jointly optimizing visual tracking and inertial pose constraints, our Inertial-monocular SLAM system based on ORB-SLAM3 [7] maintains tracking for a short period of time even if visual tracking fails due to motion blur or a lack of visual features (e.g. looking down at a table). This allows UMI to capture and deploy highly

https://umi-gripper.github.io/umi.pdf

Experiments



Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—**

***Index Terms*—bla**

I. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

II. Related Work

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

III. Method

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

IV. Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

V. Conclusion & Discussion

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

References

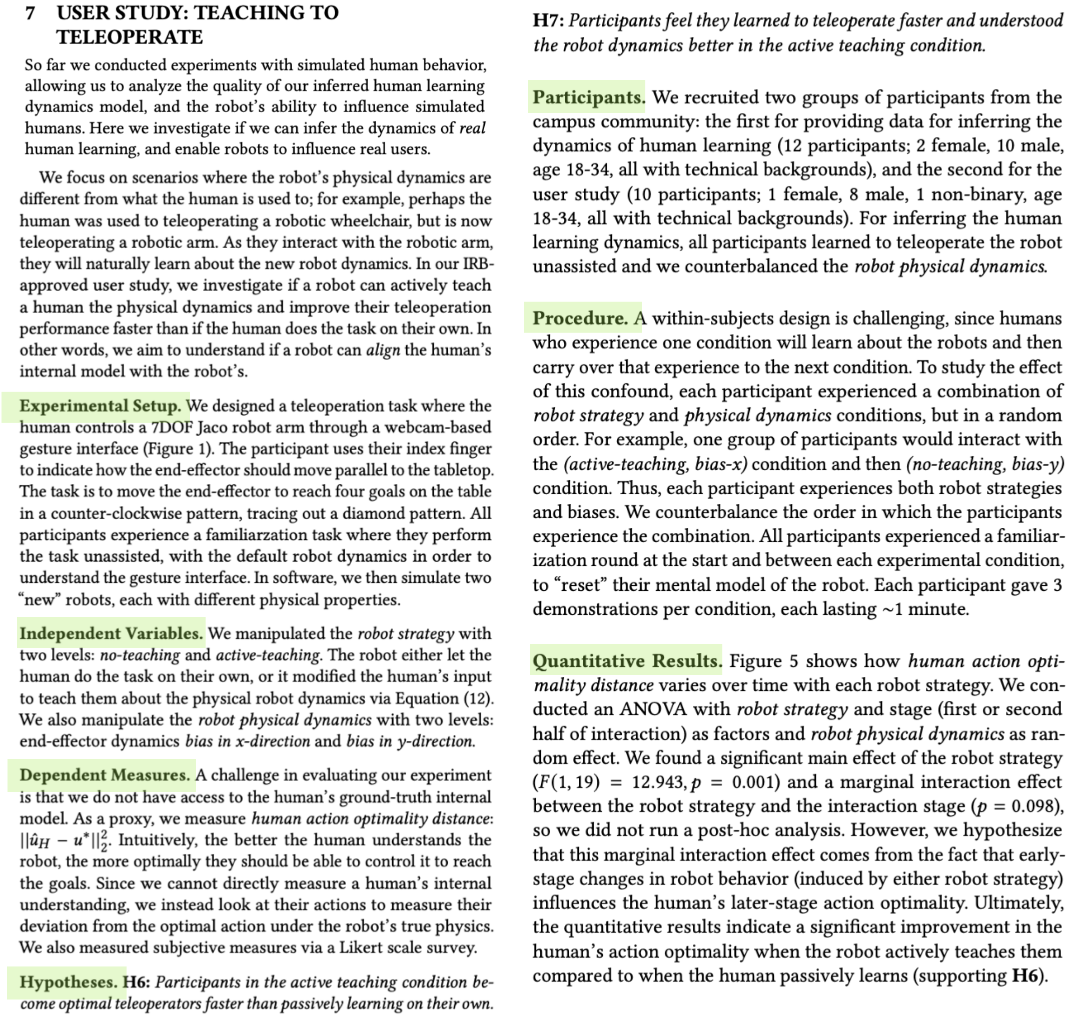
[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, "Title of paper if known," unpublished.
[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

Identify applicable funding agency here. If none, delete this.

Example 1

(user study)

Sections highlight the key things I need to know about a study



7 USER STUDY: TEACHING TO TELEOPERATE

So far we conducted experiments with simulated human behavior, allowing us to analyze the quality of our inferred human learning dynamics model, and the robot's ability to influence simulated humans. Here we investigate if we can infer the dynamics of *real* human learning, and enable robots to influence real users.

We focus on scenarios where the robot's physical dynamics are different from what the human is used to; for example, perhaps the human was used to teleoperating a robotic wheelchair, but is now teleoperating a robotic arm. As they interact with the robotic arm, they will naturally learn about the new robot dynamics. In our IRB-approved user study, we investigate if a robot can actively teach a human the physical dynamics and improve their teleoperation performance faster than if the human does the task on their own. In other words, we aim to understand if a robot can *align* the human's internal model with the robot's.

Experimental Setup. We designed a teleoperation task where the human controls a 7DOF Jaco robot arm through a webcam-based gesture interface (Figure 1). The participant uses their index finger to indicate how the end-effector should move parallel to the tabletop. The task is to move the end-effector to reach four goals on the table in a counter-clockwise pattern, tracing out a diamond pattern. All participants experience a familiarzation task where they perform the task unassisted, with the default robot dynamics in order to understand the gesture interface. In software, we then simulate two "new" robots, each with different physical properties.

Independent Variables. We manipulated the *robot strategy* with two levels: *no-teaching* and *active-teaching*. The robot either let the human do the task on their own, or it modified the human's input to teach them about the physical robot dynamics via Equation (12). We also manipulate the *robot physical dynamics* with two levels: end-effector dynamics *bias in x-direction* and *bias in y-direction*.

Dependent Measures. A challenge in evaluating our experiment is that we do not have access to the human's ground-truth internal model. As a proxy, we measure *human action optimality distance*: $||\hat{u}_H - u^*||_2^2$. Intuitively, the better the human understands the robot, the more optimally they should be able to control it to reach the goals. Since we cannot directly measure a human's internal understanding, we instead look at their actions to measure their deviation from the optimal action under the robot's true physics. We also measured subjective measures via a Likert scale survey.

Hypotheses. H6: *Participants in the active teaching condition become optimal teleoperators faster than passively learning on their own.*

H7: *Participants feel they learned to teleoperate faster and understood the robot dynamics better in the active teaching condition.*

Participants. We recruited two groups of participants from the campus community: the first for providing data for inferring the dynamics of human learning (12 participants; 2 female, 10 male, age 18-34, all with technical backgrounds), and the second for the user study (10 participants; 1 female, 8 male, 1 non-binary, age 18-34, all with technical backgrounds). For inferring the human learning dynamics, all participants learned to teleoperate the robot unassisted and we counterbalanced the *robot physical dynamics*.

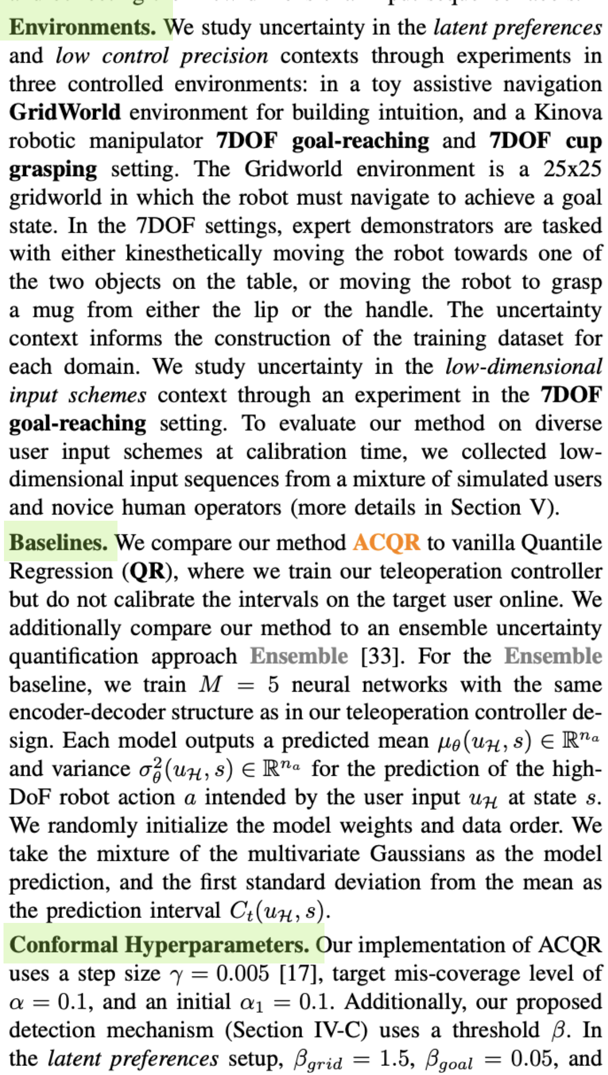
Procedure. A within-subjects design is challenging, since humans who experience one condition will learn about the robots and then carry over that experience to the next condition. To study the effect of this confound, each participant experienced a combination of *robot strategy* and *physical dynamics* conditions, but in a random order. For example, one group of participants would interact with the *(active-teaching, bias-x)* condition and then *(no-teaching, bias-y)* condition. Thus, each participant experiences both robot strategies and biases. We counterbalance the order in which the participants experience the combination. All participants experienced a familiarization round at the start and between each experimental condition, to "reset" their mental model of the robot. Each participant gave 3 demonstrations per condition, each lasting ∼1 minute.

Quantitative Results. Figure 5 shows how *human action optimality distance* varies over time with each robot strategy. We conducted an ANOVA with *robot strategy* and stage (first or second half of interaction) as factors and *robot physical dynamics* as random effect. We found a significant main effect of the robot strategy ($F(1, 19) = 12.943, p = 0.001$) and a marginal interaction effect between the robot strategy and the interaction stage ($p = 0.098$), so we did not run a post-hoc analysis. However, we hypothesize that this marginal interaction effect comes from the fact that early-stage changes in robot behavior (induced by either robot strategy) influences the human's later-stage action optimality. Ultimately, the quantitative results indicate a significant improvement in the human's action optimality when the robot actively teaches them compared to when the human passively learns (supporting **H6**).

Example 1

(simulation results)

Sections correlate with essential information I need to parse results



Environments. We study uncertainty in the *latent preferences* and *low control precision* contexts through experiments in three controlled environments: in a toy assistive navigation **GridWorld** environment for building intuition, and a Kinova robotic manipulator **7DOF goal-reaching** and **7DOF cup grasping** setting. The Gridworld environment is a 25x25 gridworld in which the robot must navigate to achieve a goal state. In the 7DOF settings, expert demonstrators are tasked with either kinesthetically moving the robot towards one of the two objects on the table, or moving the robot to grasp a mug from either the lip or the handle. The uncertainty context informs the construction of the training dataset for each domain. We study uncertainty in the *low-dimensional input schemes* context through an experiment in the **7DOF goal-reaching** setting. To evaluate our method on diverse user input schemes at calibration time, we collected low-dimensional input sequences from a mixture of simulated users and novice human operators (more details in Section V).

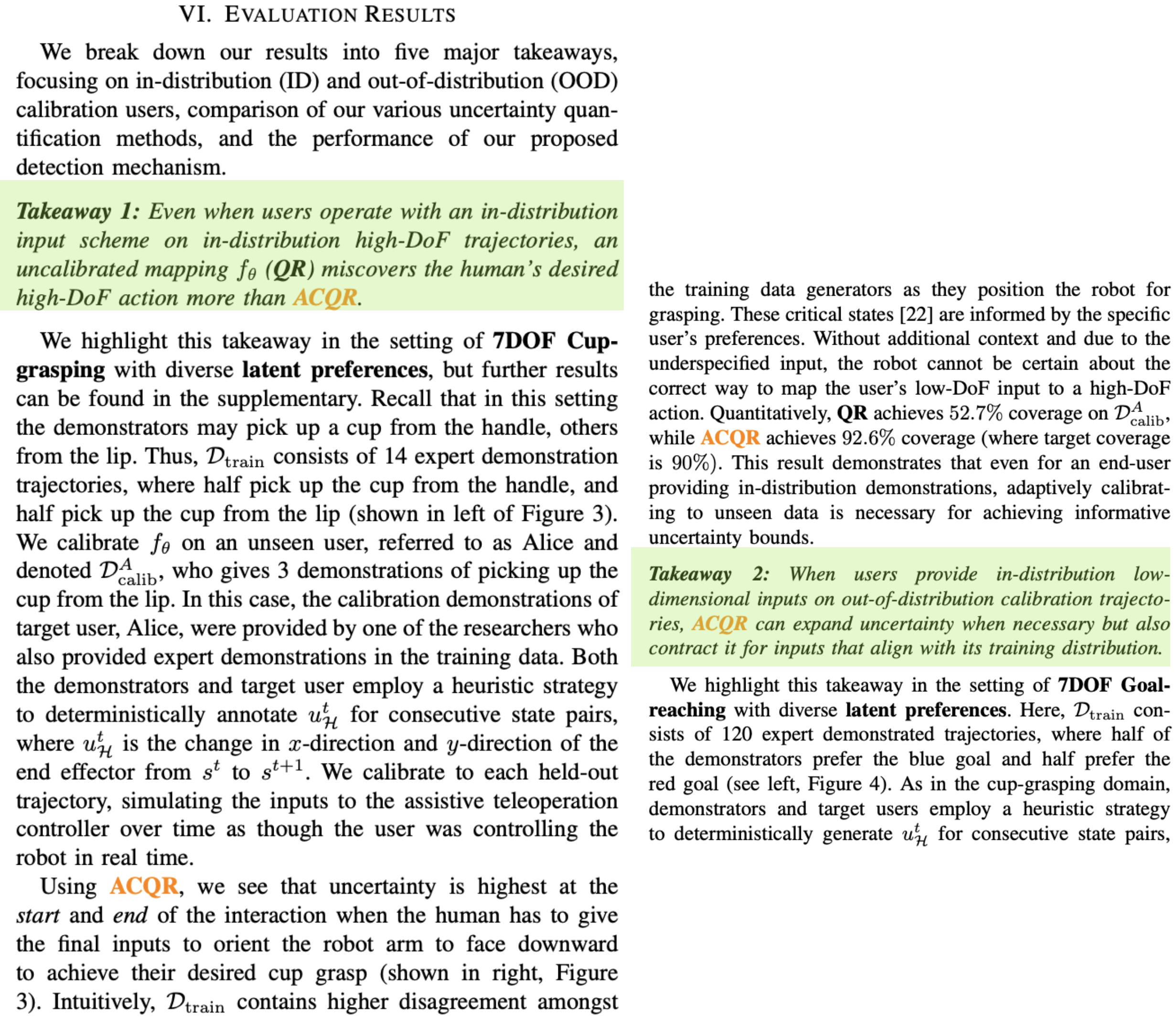
Baselines. We compare our method **ACQR** to vanilla Quantile Regression (**QR**), where we train our teleoperation controller but do not calibrate the intervals on the target user online. We additionally compare our method to an ensemble uncertainty quantification approach **Ensemble** [33]. For the **Ensemble** baseline, we train $M = 5$ neural networks with the same encoder-decoder structure as in our teleoperation controller design. Each model outputs a predicted mean $\mu_\theta(u_\mathcal{H}, s) \in \mathbb{R}^{n_a}$ and variance $\sigma^2_\theta(u_\mathcal{H}, s) \in \mathbb{R}^{n_a}$ for the prediction of the high-DoF robot action a intended by the user input $u_\mathcal{H}$ at state s. We randomly initialize the model weights and data order. We take the mixture of the multivariate Gaussians as the model prediction, and the first standard deviation from the mean as the prediction interval $C_t(u_\mathcal{H}, s)$.

Conformal Hyperparameters. Our implementation of ACQR uses a step size $\gamma = 0.005$ [17], target mis-coverage level of $\alpha = 0.1$, and an initial $\alpha_1 = 0.1$. Additionally, our proposed detection mechanism (Section IV-C) uses a threshold β. In the *latent preferences* setup, $\beta_{grid} = 1.5$, $\beta_{goal} = 0.05$, and

Example 1

(simulation results)

Takeaway blocks let the reader skim the experimental results in a very intuitive way



VI. Evaluation Results

We break down our results into five major takeaways, focusing on in-distribution (ID) and out-of-distribution (OOD) calibration users, comparison of our various uncertainty quantification methods, and the performance of our proposed detection mechanism.

Takeaway 1: *Even when users operate with an in-distribution input scheme on in-distribution high-DoF trajectories, an uncalibrated mapping f_θ (**QR**) miscovers the human's desired high-DoF action more than **ACQR**.*

We highlight this takeaway in the setting of **7DOF Cup-grasping** with diverse **latent preferences**, but further results can be found in the supplementary. Recall that in this setting the demonstrators may pick up a cup from the handle, others from the lip. Thus, $\mathcal{D}_{\text{train}}$ consists of 14 expert demonstration trajectories, where half pick up the cup from the handle, and half pick up the cup from the lip (shown in left of Figure 3). We calibrate f_θ on an unseen user, referred to as Alice and denoted $\mathcal{D}^{A}_{\text{calib}}$, who gives 3 demonstrations of picking up the cup from the lip. In this case, the calibration demonstrations of target user, Alice, were provided by one of the researchers who also provided expert demonstrations in the training data. Both the demonstrators and target user employ a heuristic strategy to deterministically annotate $u^t_{\mathcal{H}}$ for consecutive state pairs, where $u^t_{\mathcal{H}}$ is the change in x-direction and y-direction of the end effector from s^t to s^{t+1}. We calibrate to each held-out trajectory, simulating the inputs to the assistive teleoperation controller over time as though the user was controlling the robot in real time.

Using **ACQR**, we see that uncertainty is highest at the *start* and *end* of the interaction when the human has to give the final inputs to orient the robot arm to face downward to achieve their desired cup grasp (shown in right, Figure 3). Intuitively, $\mathcal{D}_{\text{train}}$ contains higher disagreement amongst the training data generators as they position the robot for grasping. These critical states [22] are informed by the specific user's preferences. Without additional context and due to the underspecified input, the robot cannot be certain about the correct way to map the user's low-DoF input to a high-DoF action. Quantitatively, **QR** achieves 52.7% coverage on $\mathcal{D}^{A}_{\text{calib}}$, while **ACQR** achieves 92.6% coverage (where target coverage is 90%). This result demonstrates that even for an end-user providing in-distribution demonstrations, adaptively calibrating to unseen data is necessary for achieving informative uncertainty bounds.

Takeaway 2: *When users provide in-distribution low-dimensional inputs on out-of-distribution calibration trajectories, **ACQR** can expand uncertainty when necessary but also contract it for inputs that align with its training distribution.*

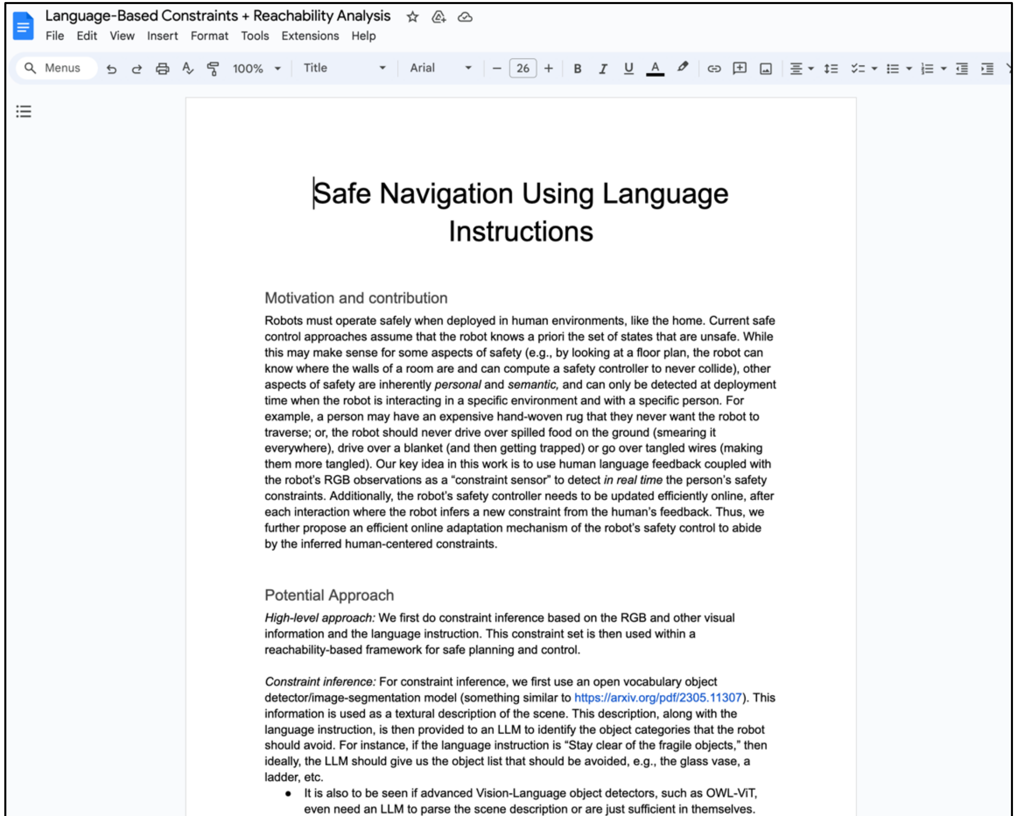
We highlight this takeaway in the setting of **7DOF Goal-reaching** with diverse **latent preferences**. Here, $\mathcal{D}_{\text{train}}$ consists of 120 expert demonstrated trajectories, where half of the demonstrators prefer the blue goal and half prefer the red goal (see left, Figure 4). As in the cup-grasping domain, demonstrators and target users employ a heuristic strategy to deterministically generate $u^t_{\mathcal{H}}$ for consecutive state pairs, It turns out all these concepts also apply to ***brainstorming*** & ***pitching research ideas***!

For example, when I think of project directions, I write down a "mock abstract" that concisely states:

- (1 sentence) What is the broad challenge
- (1-2 sentence) What is the research gap
- (1-2 sentence) Key idea of the proposed work
- (1 sentence) What is the outcome if I'm successful?

For example, when I think of project directions, I write down a "mock abstract"

Initial Project Document


Language-Based Constraints + Reachability Analysis
File Edit View Insert Format Tools Extensions Help
Menus
100%
Title
Arial
26

Safe Navigation Using Language Instructions

Motivation and contribution

Robots must operate safely when deployed in human environments, like the home. Current safe control approaches assume that the robot knows a priori the set of states that are unsafe. While this may make sense for some aspects of safety (e.g., by looking at a floor plan, the robot can know where the walls of a room are and can compute a safety controller to never collide), other aspects of safety are inherently *personal* and *semantic,* and can only be detected at deployment time when the robot is interacting in a specific environment and with a specific person. For example, a person may have an expensive hand-woven rug that they never want the robot to traverse; or, the robot should never drive over spilled food on the ground (smearing it everywhere), drive over a blanket (and then getting trapped) or go over tangled wires (making them more tangled). Our key idea in this work is to use human language feedback coupled with the robot's RGB observations as a "constraint sensor" to detect *in real time* the person's safety constraints. Additionally, the robot's safety controller needs to be updated efficiently online, after each interaction where the robot infers a new constraint from the human's feedback. Thus, we further propose an efficient online adaptation mechanism of the robot's safety control to abide by the inferred human-centered constraints.

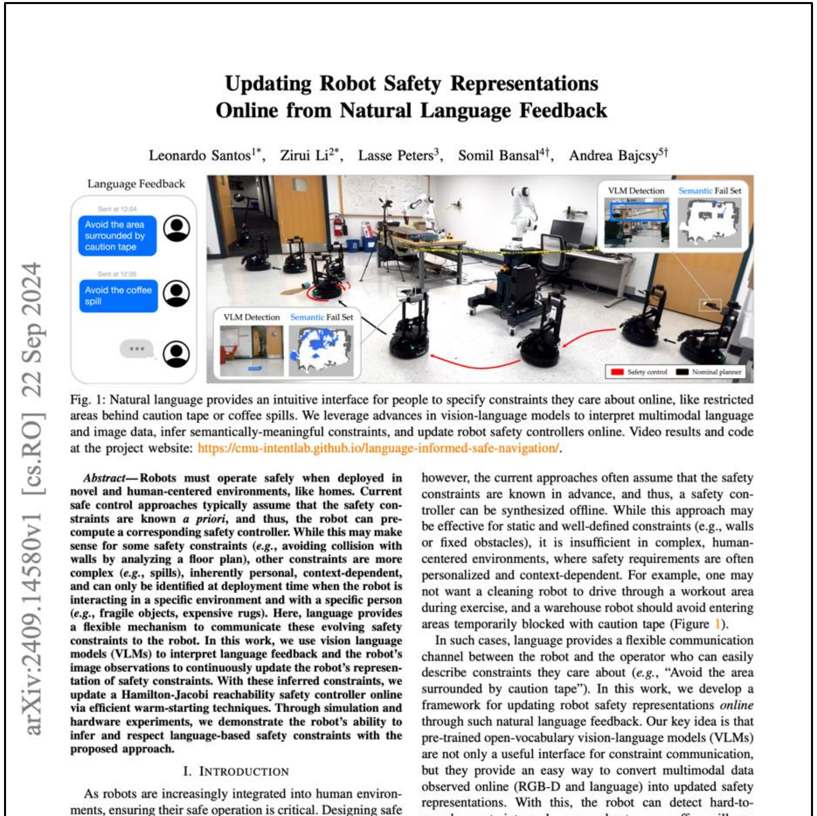
Potential Approach

High-level approach: We first do constraint inference based on the RGB and other visual information and the language instruction. This constraint set is then used within a reachability-based framework for safe planning and control.

Constraint inference: For constraint inference, we first use an open vocabulary object detector/image-segmentation model (something similar to https://arxiv.org/pdf/2305.11307). This information is used as a textural description of the scene. This description, along with the language instruction, is then provided to an LLM to identify the object categories that the robot should avoid. For instance, if the language instruction is "Stay clear of the fragile objects," then ideally, the LLM should give us the object list that should be avoided, e.g., the glass vase, a ladder, etc.

- It is also to be seen if advanced Vision-Language object detectors, such as OWL-ViT, even need an LLM to parse the scene description or are just sufficient in themselves.

Final Paper



Updating Robot Safety Representations Online from Natural Language Feedback

Leonardo Santos[1*], Zirui Li[2*], Lasse Peters[3], Somil Bansal[4†], Andrea Bajcsy[5†]

Fig. 1: Natural language provides an intuitive interface for people to specify constraints they care about online, like restricted areas behind caution tape or coffee spills. We leverage advances in vision-language models to interpret multimodal language and image data, infer semantically-meaningful constraints, and update robot safety controllers online. Video results and code at the project website: https://cmu-intentlab.github.io/language-informed-safe-navigation/.

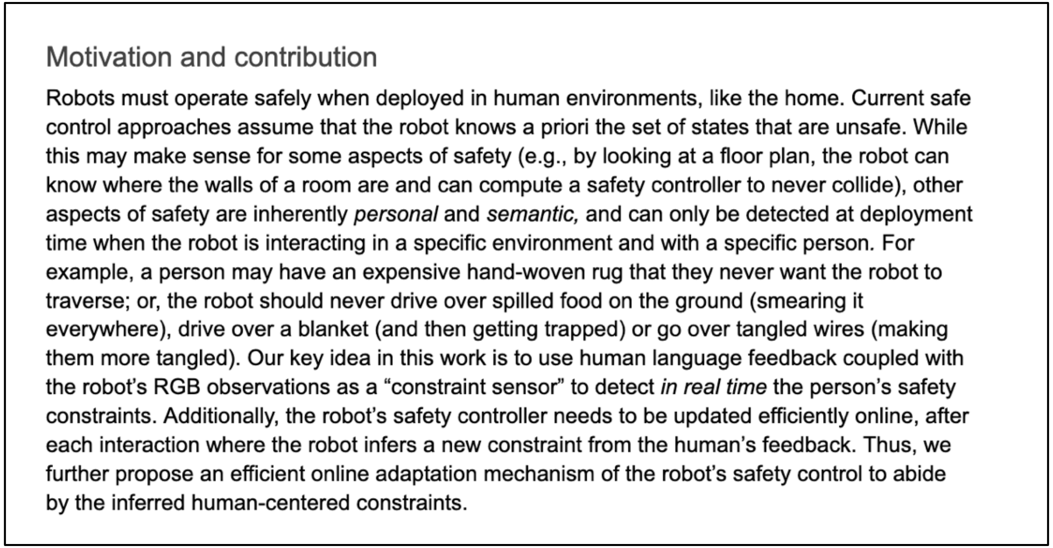
***Abstract*— Robots must operate safely when deployed in novel and human-centered environments, like homes. Current safe control approaches typically assume that the safety constraints are known *a priori*, and thus, the robot can pre-compute a corresponding safety controller. While this may make sense for some safety constraints (*e.g.*, avoiding collision with walls by analyzing a floor plan), other constraints are more complex (*e.g.*, spills), inherently personal, context-dependent, and can only be identified at deployment time when the robot is interacting in a specific environment and with a specific person (*e.g.*, fragile objects, expensive rugs). Here, language provides a flexible mechanism to communicate these evolving safety constraints to the robot. In this work, we use vision language models (VLMs) to interpret language feedback and the robot's image observations to continuously update the robot's representation of safety constraints. With these inferred constraints, we update a Hamilton-Jacobi reachability safety controller online via efficient warm-starting techniques. Through simulation and hardware experiments, we demonstrate the robot's ability to infer and respect language-based safety constraints with the proposed approach.**

I. Introduction

As robots are increasingly integrated into human environments, ensuring their safe operation is critical. Designing safe however, the current approaches often assume that the safety constraints are known in advance, and thus, a safety controller can be synthesized offline. While this approach may be effective for static and well-defined constraints (e.g., walls or fixed obstacles), it is insufficient in complex, human-centered environments, where safety requirements are often personalized and context-dependent. For example, one may not want a cleaning robot to drive through a workout area during exercise, and a warehouse robot should avoid entering areas temporarily blocked with caution tape (Figure 1).

In such cases, language provides a flexible communication channel between the robot and the operator who can easily describe constraints they care about (*e.g.*, "Avoid the area surrounded by caution tape"). In this work, we develop a framework for updating robot safety representations *online* through such natural language feedback. Our key idea is that pre-trained open-vocabulary vision-language models (VLMs) are not only a useful interface for constraint communication, but they provide an easy way to convert multimodal data observed online (RGB-D and language) into updated safety representations. With this, the robot can detect hard-to-

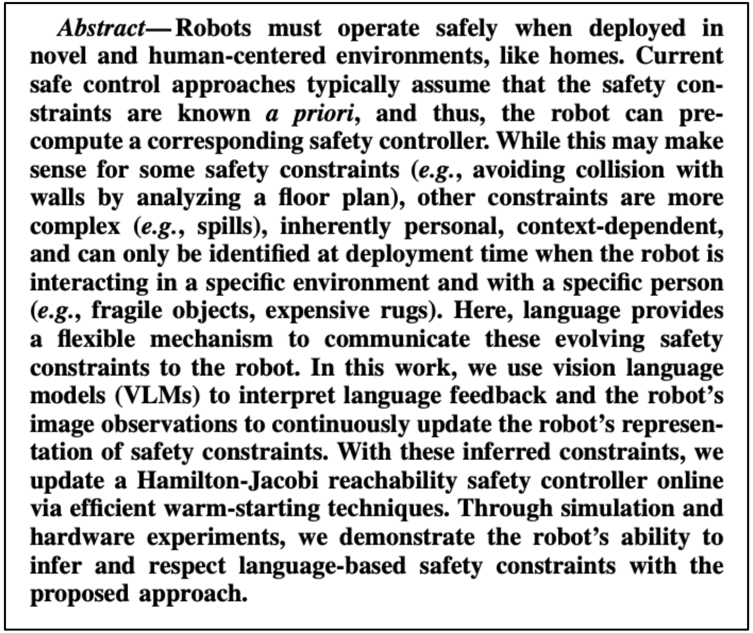
Initial Project Document



Motivation and contribution

Robots must operate safely when deployed in human environments, like the home. Current safe control approaches assume that the robot knows a priori the set of states that are unsafe. While this may make sense for some aspects of safety (e.g., by looking at a floor plan, the robot can know where the walls of a room are and can compute a safety controller to never collide), other aspects of safety are inherently *personal* and *semantic,* and can only be detected at deployment time when the robot is interacting in a specific environment and with a specific person. For example, a person may have an expensive hand-woven rug that they never want the robot to traverse; or, the robot should never drive over spilled food on the ground (smearing it everywhere), drive over a blanket (and then getting trapped) or go over tangled wires (making them more tangled). Our key idea in this work is to use human language feedback coupled with the robot's RGB observations as a "constraint sensor" to detect *in real time* the person's safety constraints. Additionally, the robot's safety controller needs to be updated efficiently online, after each interaction where the robot infers a new constraint from the human's feedback. Thus, we further propose an efficient online adaptation mechanism of the robot's safety control to abide by the inferred human-centered constraints.

Final Paper



***Abstract*— Robots must operate safely when deployed in novel and human-centered environments, like homes. Current safe control approaches typically assume that the safety constraints are known *a priori*, and thus, the robot can precompute a corresponding safety controller. While this may make sense for some safety constraints (*e.g.*, avoiding collision with walls by analyzing a floor plan), other constraints are more complex (*e.g.*, spills), inherently personal, context-dependent, and can only be identified at deployment time when the robot is interacting in a specific environment and with a specific person (*e.g.*, fragile objects, expensive rugs). Here, language provides a flexible mechanism to communicate these evolving safety constraints to the robot. In this work, we use vision language models (VLMs) to interpret language feedback and the robot's image observations to continuously update the robot's representation of safety constraints. With these inferred constraints, we update a Hamilton-Jacobi reachability safety controller online via efficient warm-starting techniques. Through simulation and hardware experiments, we demonstrate the robot's ability to infer and respect language-based safety constraints with the proposed approach.**

Most common failure modes:

1) failure to *distill key idea*
2) poor argument *structure*

writer
(i.e., you)

Most common failure modes:

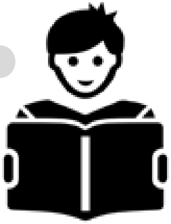
1) failure to *distill key idea*
2) poor argument *structure*

What is the contribution?

Why is this new?

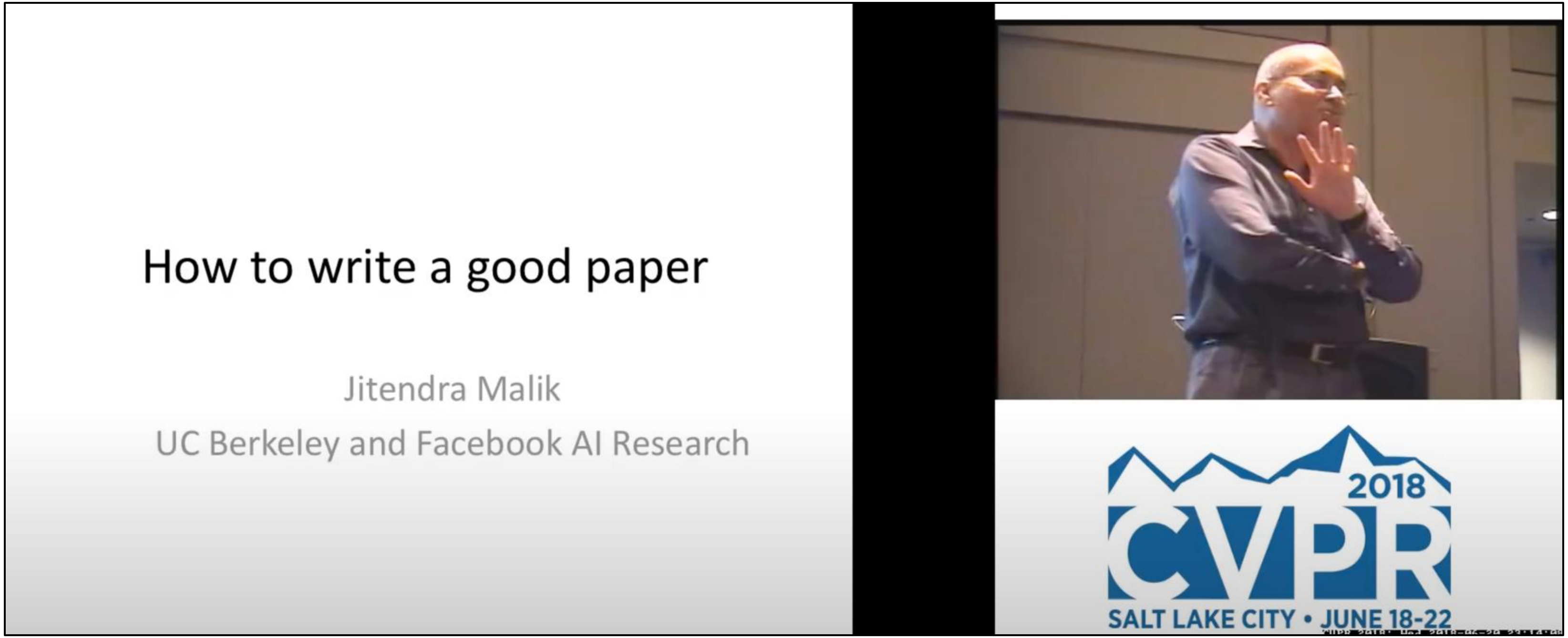
Why did you try X and not Y?

writer
(i.e., you)



reader
(i.e., Reviewer 2)

More resources



How to write a good paper

Jitendra Malik

UC Berkeley and Facebook AI Research

https://www.youtube.com/watch?v=imEtTnQKt4M

16-886 Embodied AI Safety

Research Skills

Technical writing

Clear writing is clear thinking.

Work *coarse to fine-grained* when writing papers. You are a sculptor!

- Distill + emphasize key ideas and logical arguments.
- Make it easy for reader to pay attention to the right stuff

Carnegie Mellon University

Andrea Bajcsy
abajcsy@cmu.edu

intent
ROBOTICS LAB