Rules of the Road: Formal Guarantees for Autonomous Vehicles with Behavioral Contract Design

Karena X. Cai*, Tung Phan-Minh*, PIs: Richard M. Murray, Soon-Jo Chung, California Institute of Technology
https://vehical.org

Abstract:

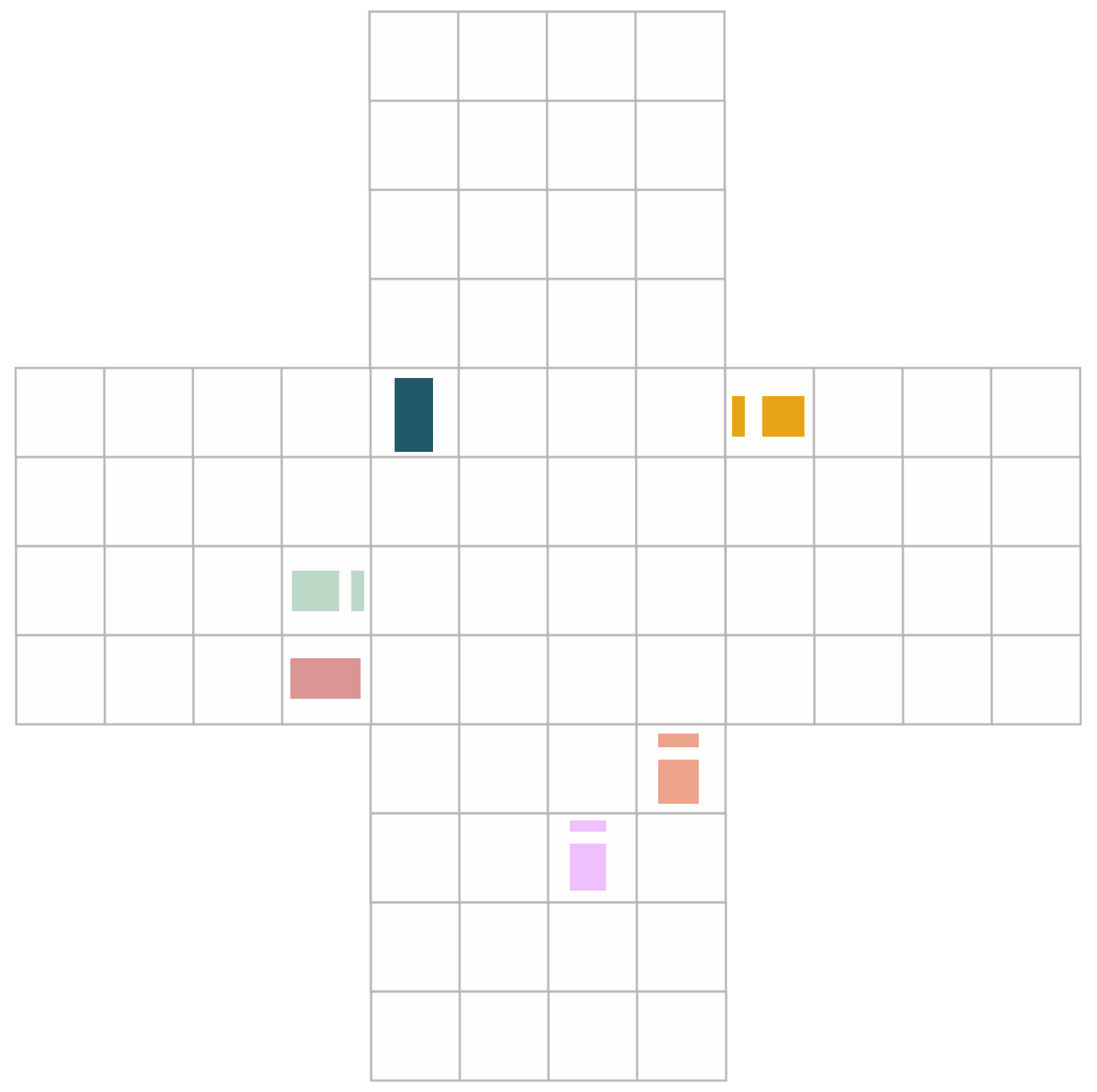
The ability to make formal guarantees on safety and performance for autonomous vehicles in **highly-interactive, dense environments** largely remains unsolved. With a **well-defined behavioral contract**, we can not only provide **formal guarantees** on agent safety and progress, but we also have a mechanism for **assigning blame** when accidents invariably occur. In this paper, we define a behavioral contract for a particular class of agents on a **road network environment** in a **quasi-simultaneous discrete-time game**. We provide **proofs of correctness** of the behavioral contract and **validate our results** in simulation.

Challenge:

How do we design a high-level decision making strategy for autonomous agents in highly-interactive environments to behave 'correctly', i.e. be safe, be lawful, and make progress towards its destination?

Extremely challenging because:

- Robot-freezing problem and unbounded rationality.
- Joint action space grows exponentially.
- Other agents can act to intentionally make safety impossible.
- Can't satisfy all road rules all the time, which to violate?



Scientific Impact:

Agent strategy (defined in a discrete-game and in specific road network environments that provides:

Safety guarantee

Safety Theorem

Given that all agents $Ag \in \mathfrak{A}$ in the quasi-simultaneous game $\mathfrak{G}$ select actions in accordance to the agent protocol defined, we can show the safety property:

$P \implies \Box Q$

P assertion that the game is in a state where every agent has a backup plan action that is safe.

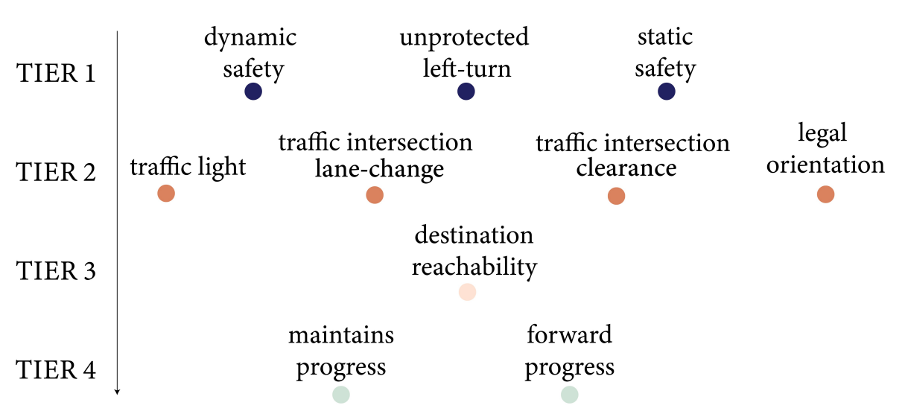
Qassertion that agents never occupy the same grid point at the same time.

Performance guarantee

Liveness Theorem

Given the sparsity conditions hold, and that all agents $Ag \in \mathfrak{A}$ in the quasi-simultaneous game $\mathfrak{G}$ select actions in accordance to the agent protocol defined, we can show all agents will eventually reach their respective destinations.

Scalability & Interpretability



Notion of Blame/Liability

$$C_j = (A_j, G_j)$$
$$\forall j \in \mathcal{J}. \forall i \in \mathcal{J} - j. G_j \subseteq A_i$$

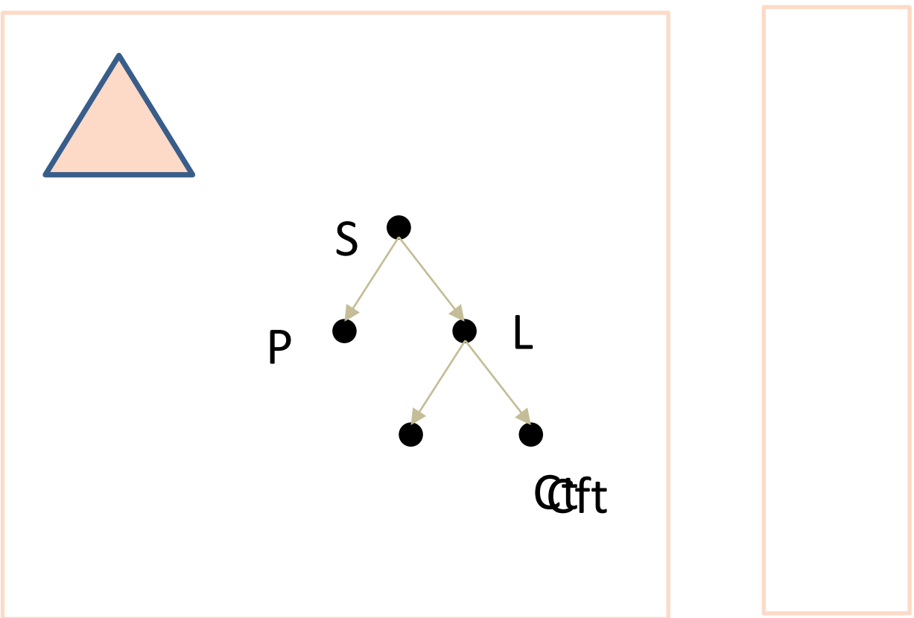
Definition II.2 (Blameworthy action)**.** *A blameworthy action/strategy is one in which an agent violates its guarantees, thereby causing another agent's assumptions not to be satisfied and thus resulting in an unwanted situation where blame must be assigned.*

Solution:

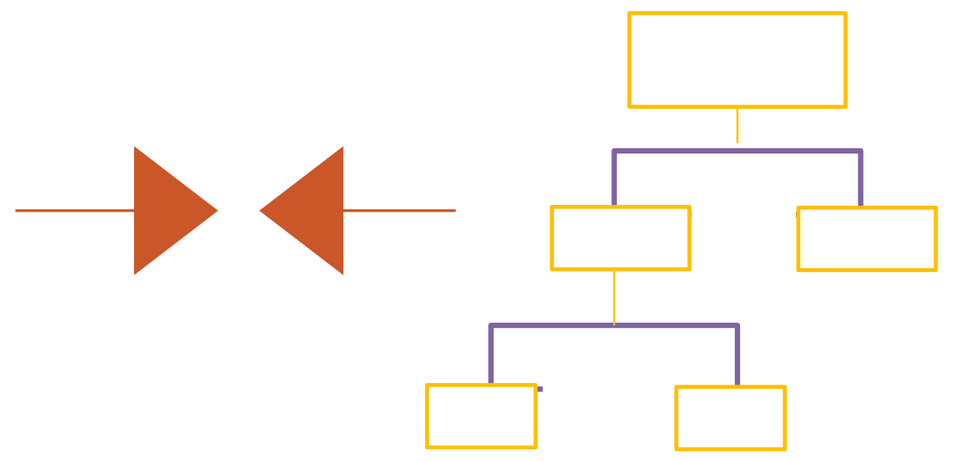
Propose the design of a behavioral protocol agents should use to select actions.

Strategy ensures agents are always entitled to safely execute their backup plan action (i.e. maximal braking)

Pt. 1 Behavioral Profile



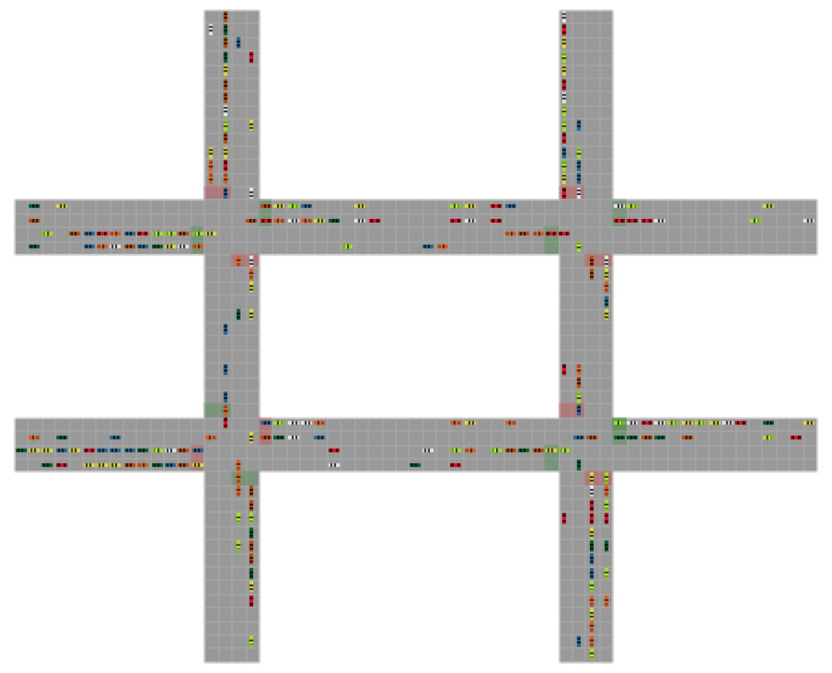
Pt. 2 Conflict Resolution Scheme



Proofs

1. Safety: no collisions.
2. Performance: agents make progress towards destinations. (under sparsity assumptions)

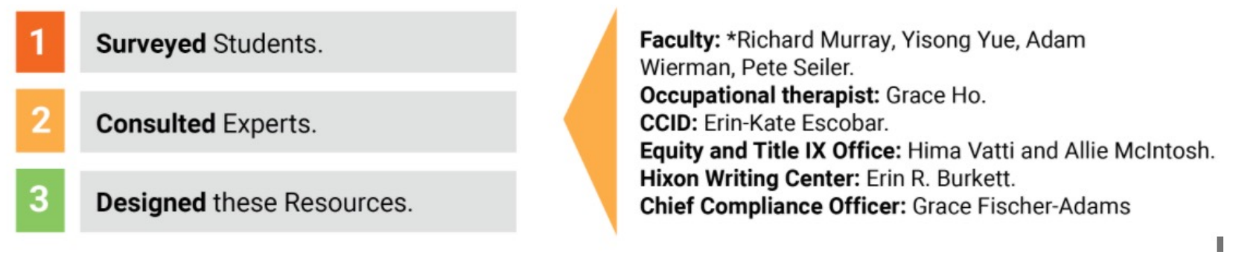
Simulations



Broader Impact on Society

- Adoption of this type of framework will lead to safer and more interpretable autonomous vehicles on the road..
- Serves as a novel framework for designing vehicle behavior with the collective in mind (instead of the individual).
- Could be integrated alongside data-driven/machine learning approaches.

Broader Impact: Education and Outreach



1 **Surveyed** Students.

2 **Consulted** Experts.

3 **Designed** these Resources.

Faculty: *Richard Murray, Yisong Yue, Adam Wierman, Pete Seiler.
Occupational therapist: Grace Ho.
CCID: Erin-Kate Escobar.
Equity and Title IX Office: Hima Vatti and Allie McIntosh.
Hixon Writing Center: Erin R. Burkett.
Chief Compliance Officer: Grace Fischer-Adams

Designed and hosted workshop on 'Building Effective Research Collaborations' to teach grad students communication and conflict prevention/management skills. Resources can be found: http://healthycollab.caltech.edu/

Quantifying Broader Impact:

- Potential to design autonomous vehicle algorithms that reduce number of collisions on the road.
- Also could help inform design of autonomous vehicle road rules and regulations.

2021 NSF Cyber-Physical Systems Principal Investigators' Meeting
June 2-4, 2021

Award ID#: CNS-1545126