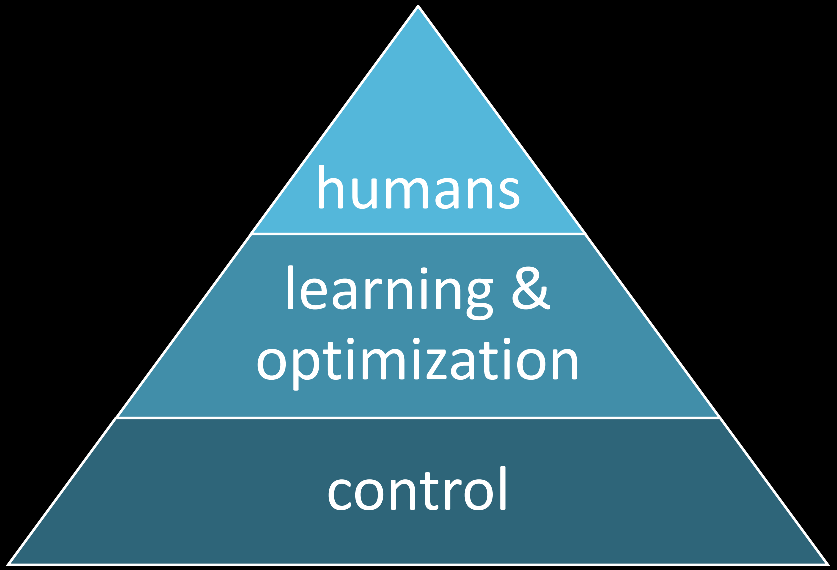
Trustworthy Autonomy:
Behavior Prediction and Validation

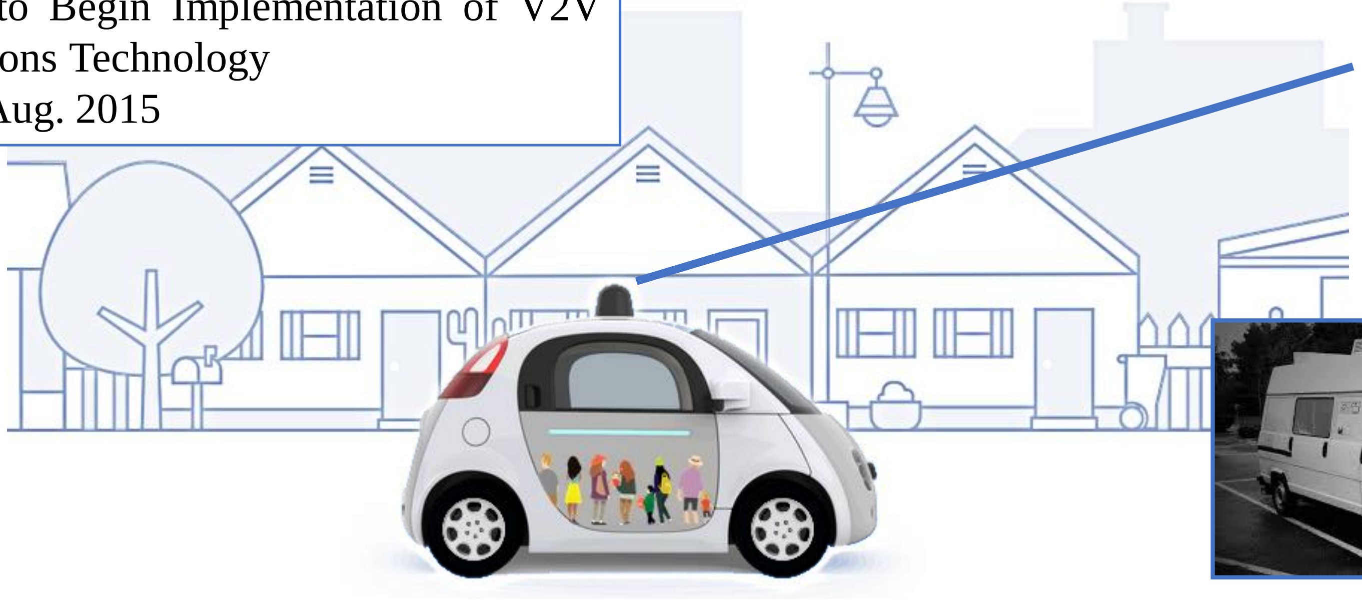
Katie Driggs-Campbell
Department of Electrical and Computer Engineering
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

How can we ensure safety in data-driven robotic systems that operate with people in the real-world?

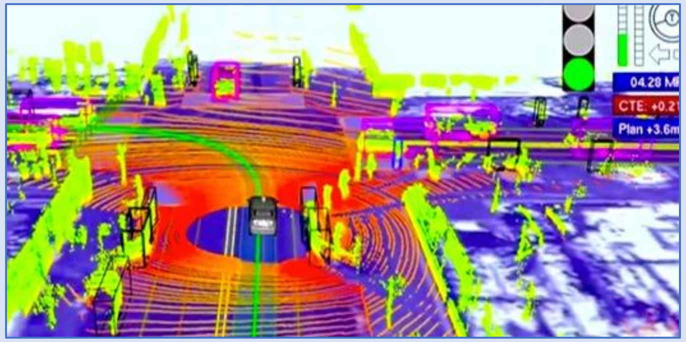


Cars are *communicating* more and more

USDOT Issues Advance Notice of Proposed Rulemaking to Begin Implementation of V2V Communications Technology
— NHTSA, Aug. 2015



Cars are *sensing* more and more



driverless cars date back to the 80s/90s in the Eureka/ Prometheus Project

There is a greater societal push than ever before…

THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate

BUSINESS | AUTOS & TRANSPORTATION | AUTOS

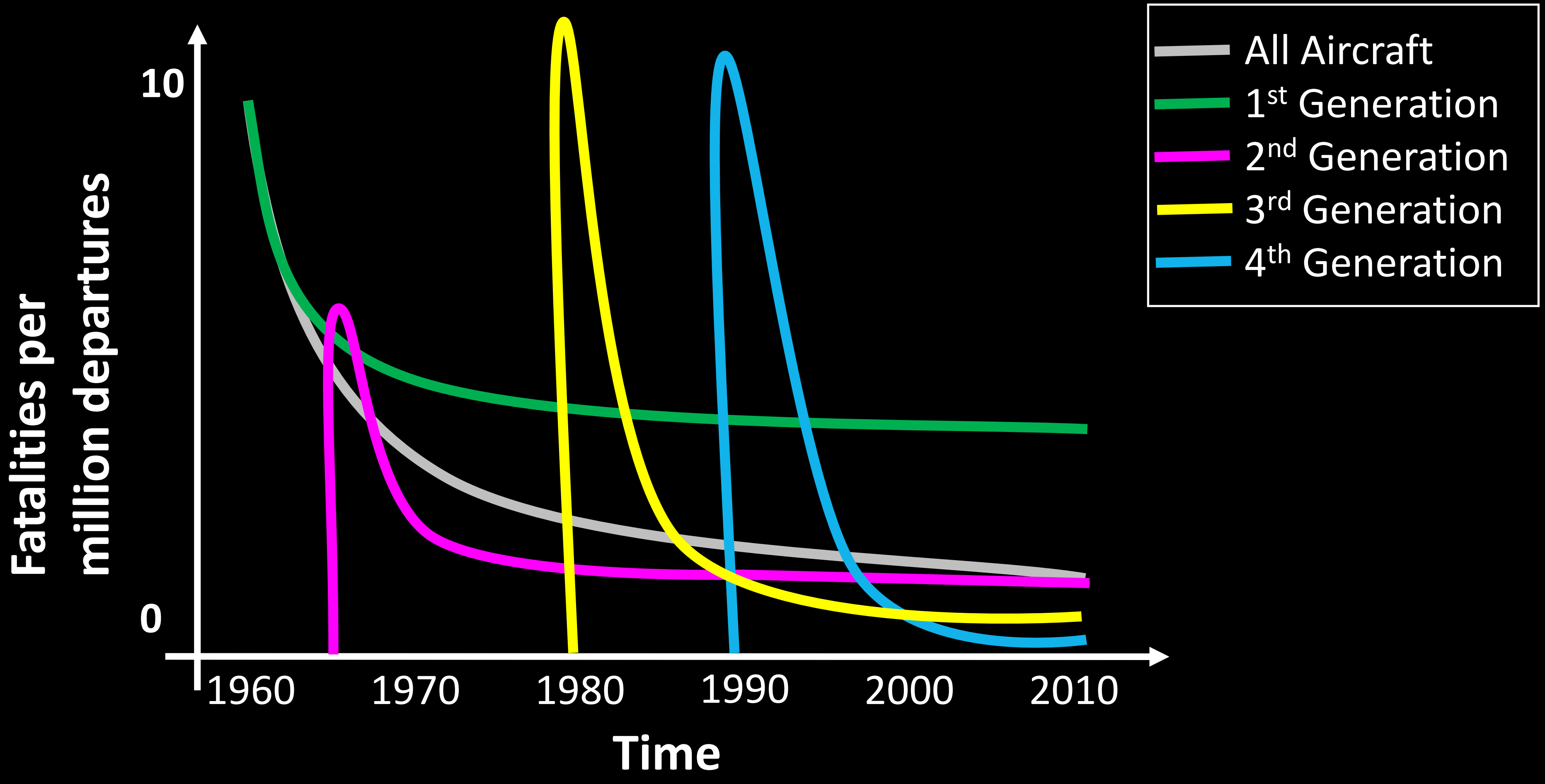
U.S. Proposes Spending $4 Billion to Encourage Driverless Cars

Obama administration aims to remove hurdles to making autonomous cars more widespread

Autonomous Vehicles in the News

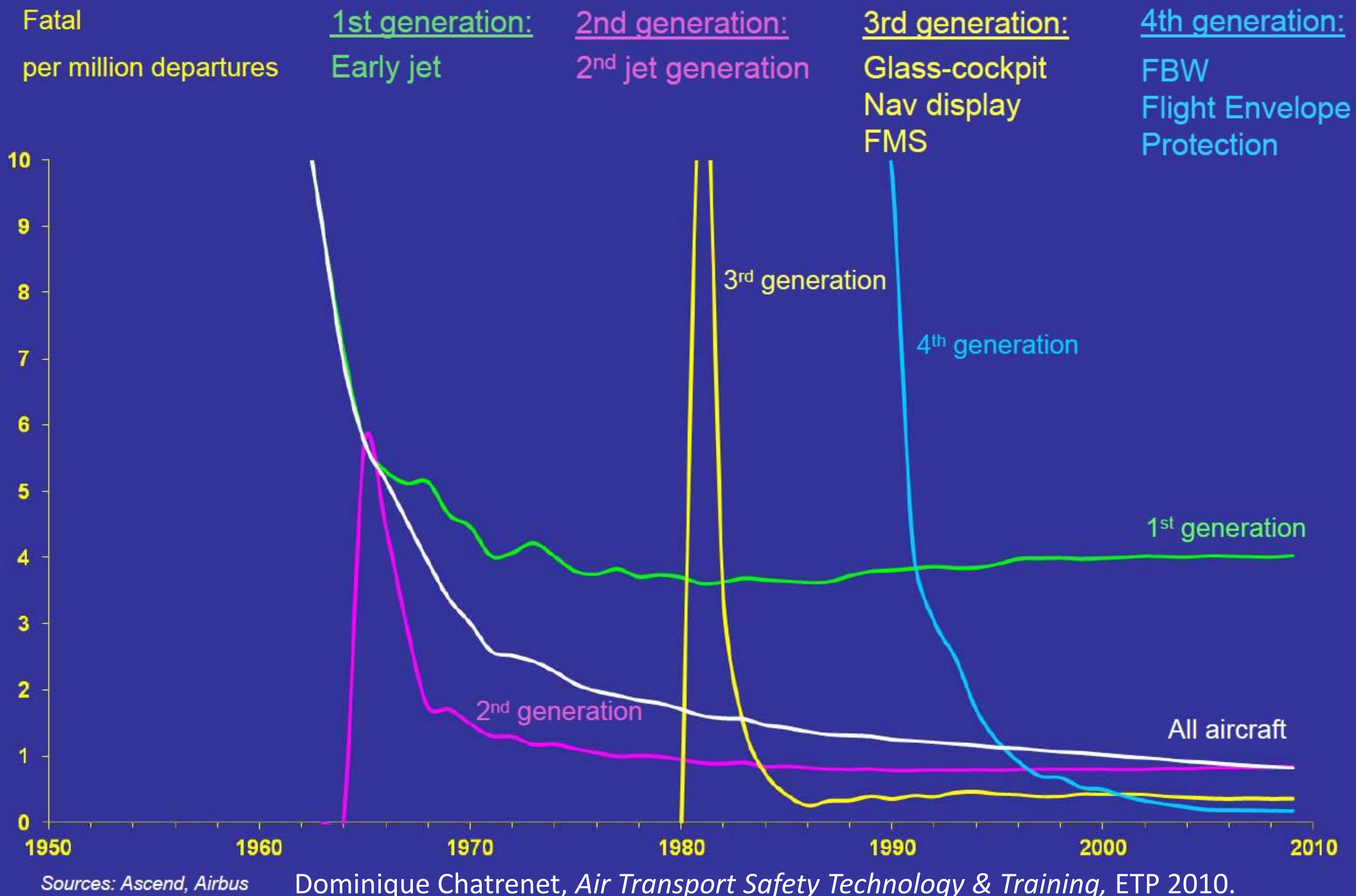
Steven Crowe, *How California's Self-Driving Cars Performed in 2017*, The Robot Report, February 2018.

Emergence of Autonomy in Planes



Dominique Chatrenet, *Air Transport Safety Technology & Training,* ETP 2010.

Fatal rate by year - valid end 2009



Dominique Chatrenet, *Air Transport Safety Technology & Training,* ETP 2010.

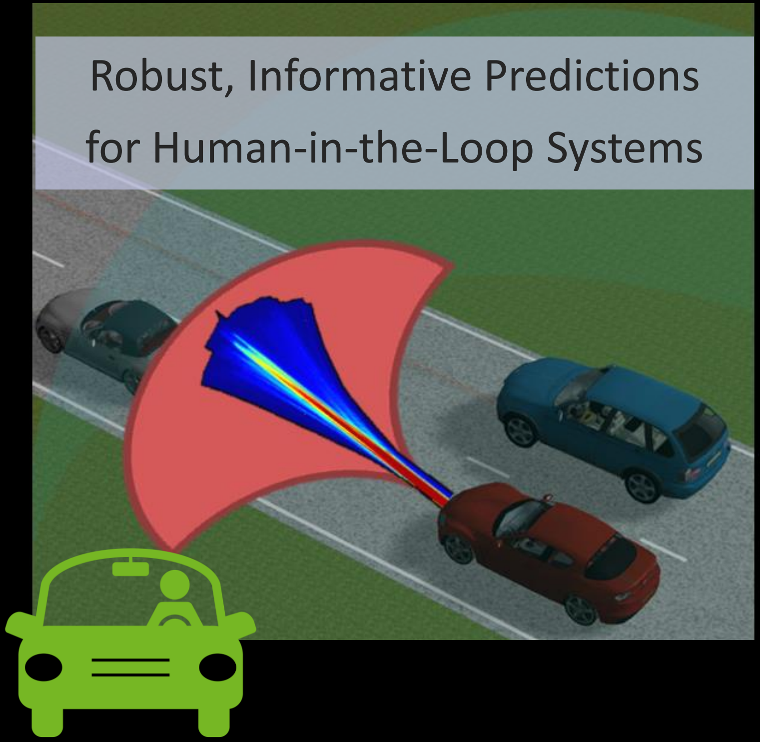
Human-Centered Autonomy

Modeling, Semi-autonomy,
& Experimental Testbeds

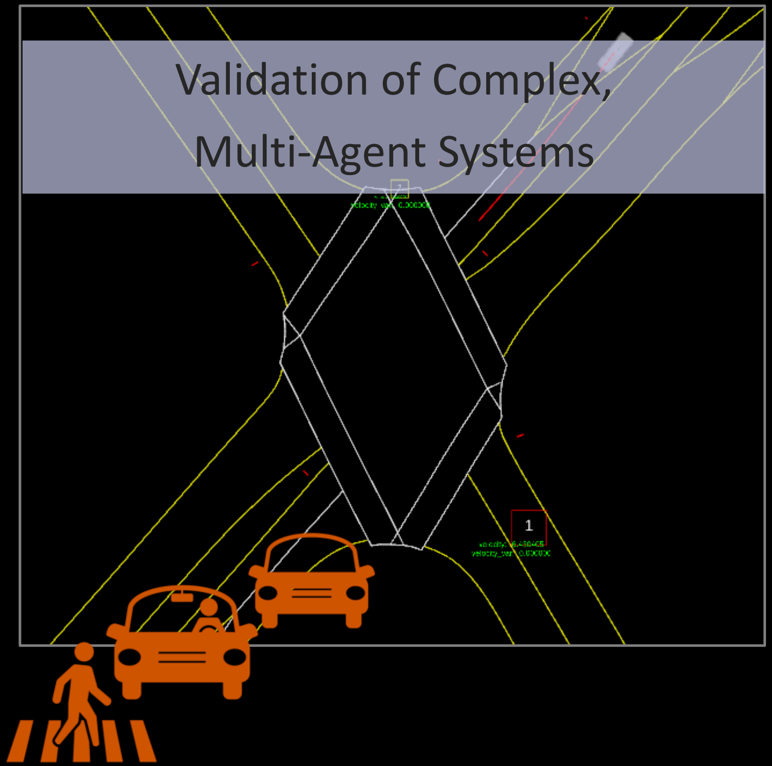
Autonomous Planning & Control

Multi-Agent Perception
& Decision Making

Roadmap

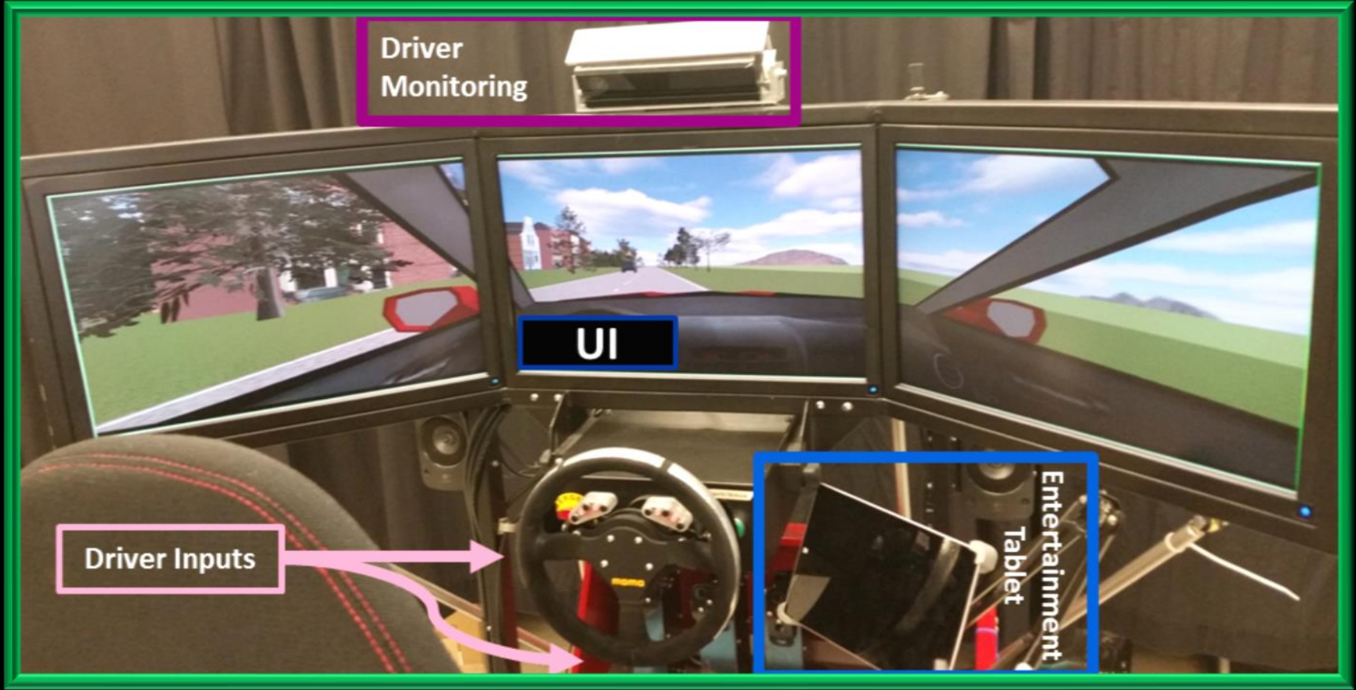

Robust, Informative Predictions
for Human-in-the-Loop Systems

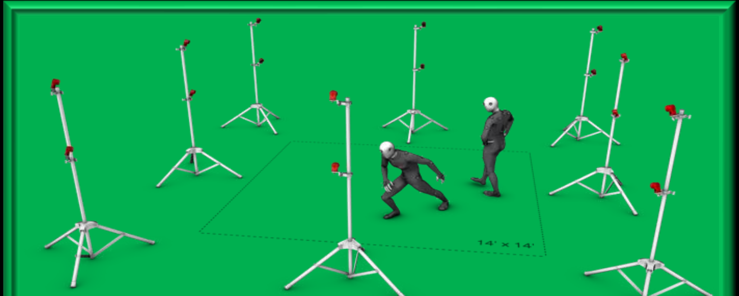

Towards Safe and
Efficient Urban Driving


Validation of Complex,
Multi-Agent Systems
1

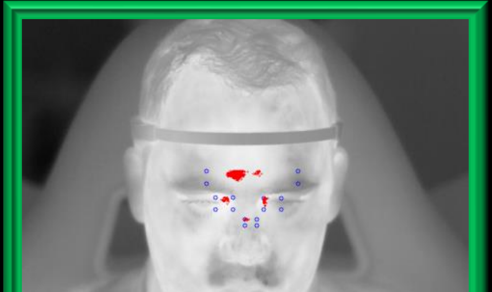
K. Driggs-Campbell, *Experimental Design for Human-in-the-Loop Driving Simulations*,
Master’s Thesis. EECS Department, University of California, Berkeley, 2015.

Outfitted for Driver Monitoring



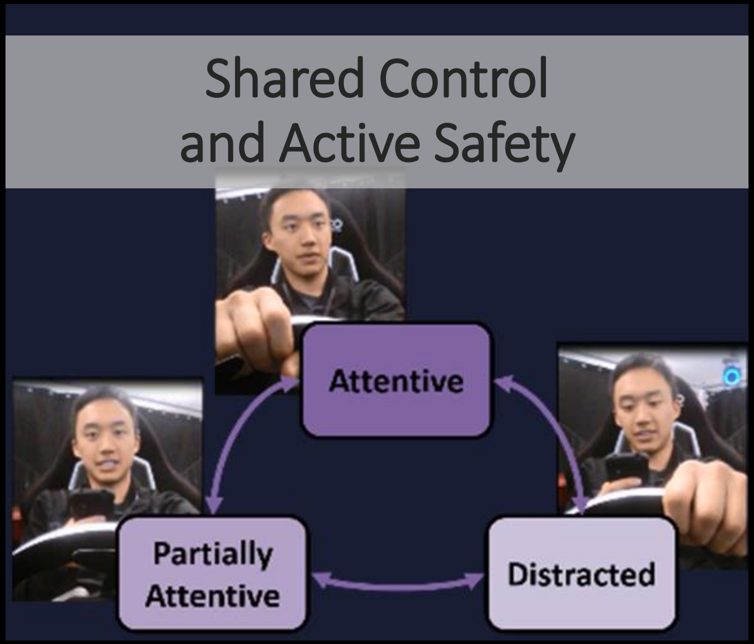

Motion Capture System

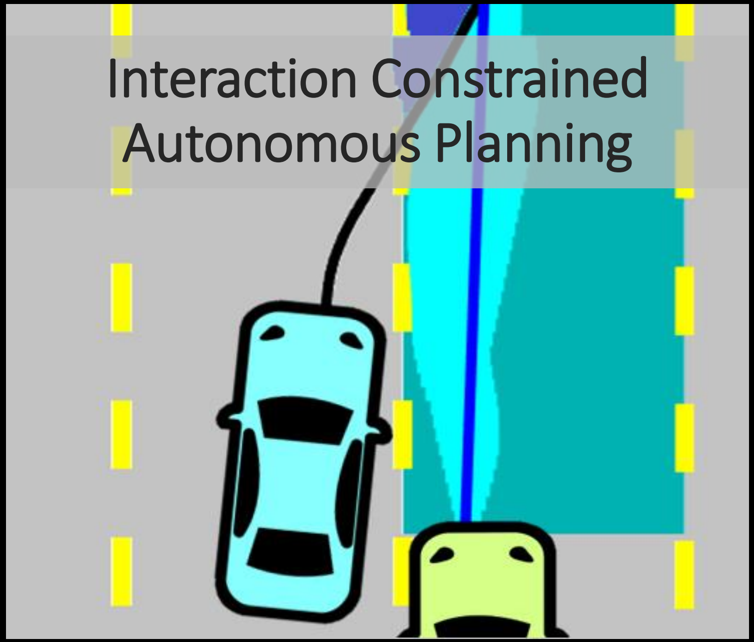

Eye Tracking

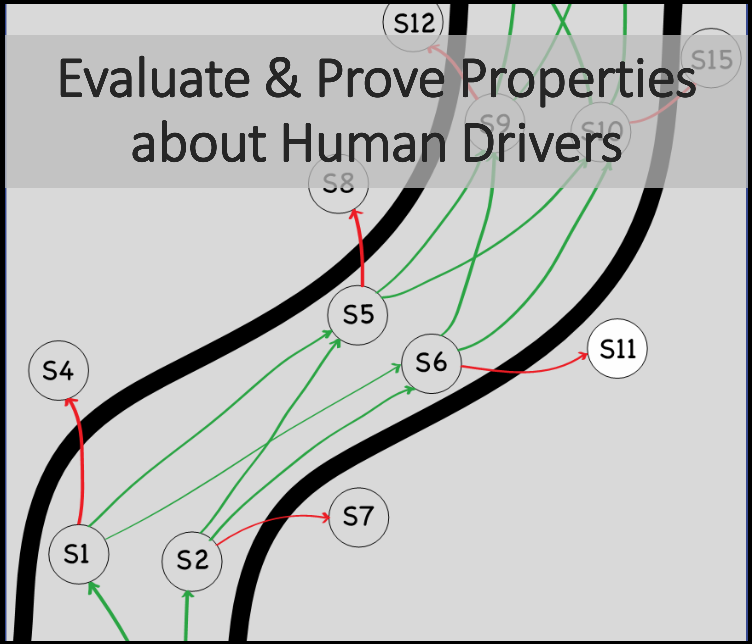

Affective Sensing

K. Driggs-Campbell, *Experimental Design for Human-in-the-Loop Driving Simulations*, Master's Thesis. EECS Department, University of California, Berkeley, 2015.

Applications for Driver Modeling





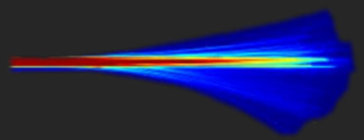


K. Driggs-Campbell, et al., *Improved Driver Modeling for Human-in-the-Loop Control*, ICRA 2015.
V. Shia, Y. Gao, R. Vasudevan, **K. Driggs-Campbell,** et al. *Semi-Autonomous Vehicular Control Using Driver Modeling*, in Transactions on ITS 2014.
K. Driggs-Campbell, et al., *Integrating Intuitive Driver Models in Autonomous Planning for Interactive Maneuvers,* in Transactions on ITS 2017.
D. Sadigh, **K. Driggs-Campbell**, et al. *Data-driven probabilistic modeling and verification of human driver behavior*, in AAAI 2014.

Driver Modeling and Active Safety

If we can identify the driver state and effectively predict their likely behavior, can we design better, less invasive active safety systems?

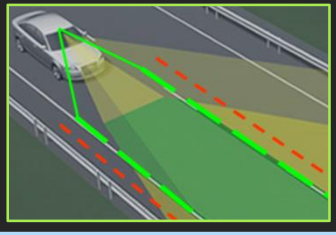
Driver Prediction



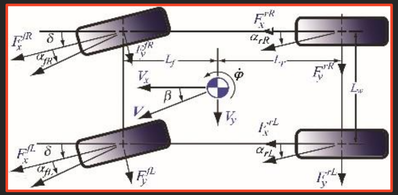
Intervention Function

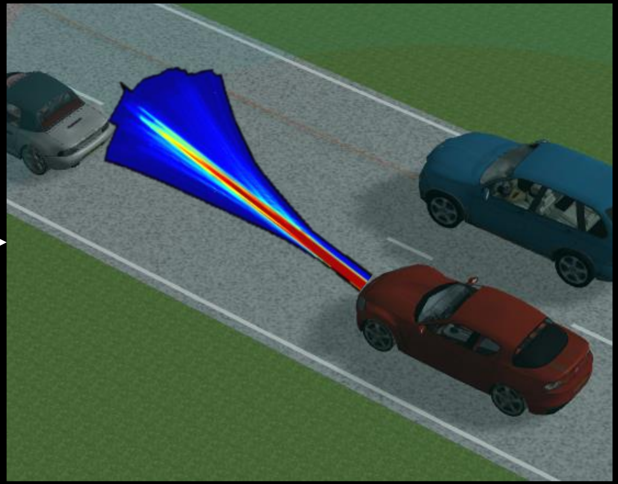
$$\mathcal{G}_k(\alpha,\tau) = \begin{cases} 1, \text{if } \exists k \ s.t. |\Delta_k \cap \mathcal{C}_k| \geq \tau \\ 0, \text{otherwise} \end{cases}$$

Environment Model



Model Predictive Control

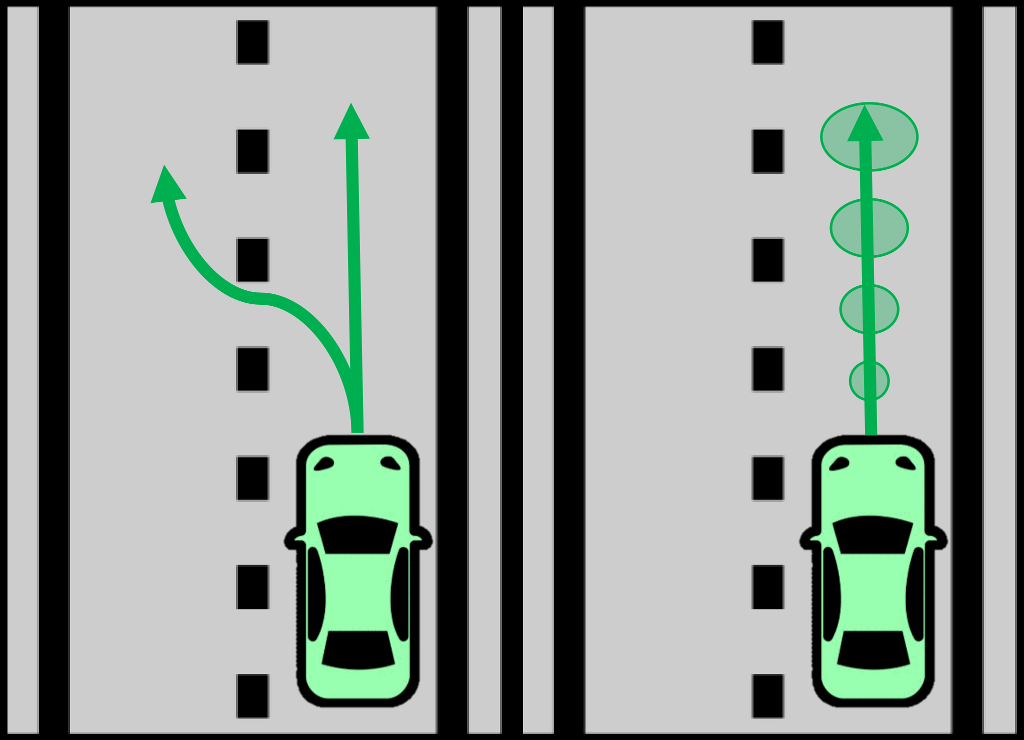




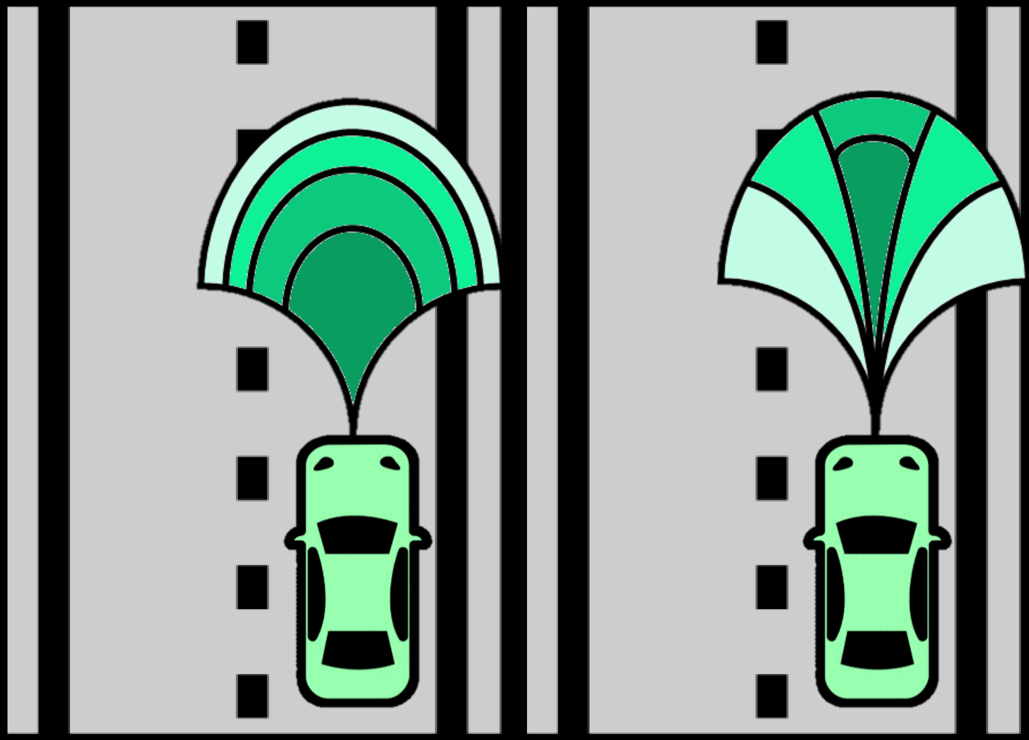
K. Driggs-Campbell, et al., *Improved Driver Modeling for Human-in-the-Loop Control*, ICRA 2015.
V. Shia, Y. Gao, R. Vasudevan, **K. Driggs-Campbell,** et al. *Semi-Autonomous Vehicular Control Using Driver Modeling*, Transactions on ITS 2014.

Predictive Modeling

Informative Models



Robust Models



Empirical Reachable Sets

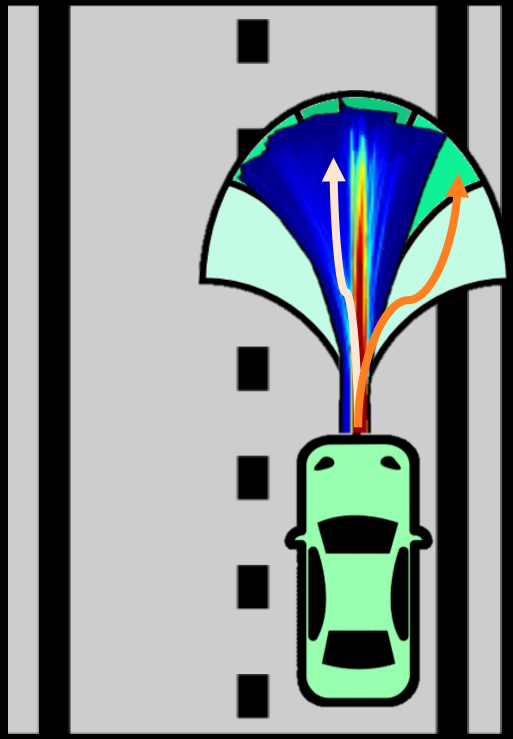
Approximate stochastic reachability with an *empirical reachable set,* by:

maximizing precision

while maintaining accuracy

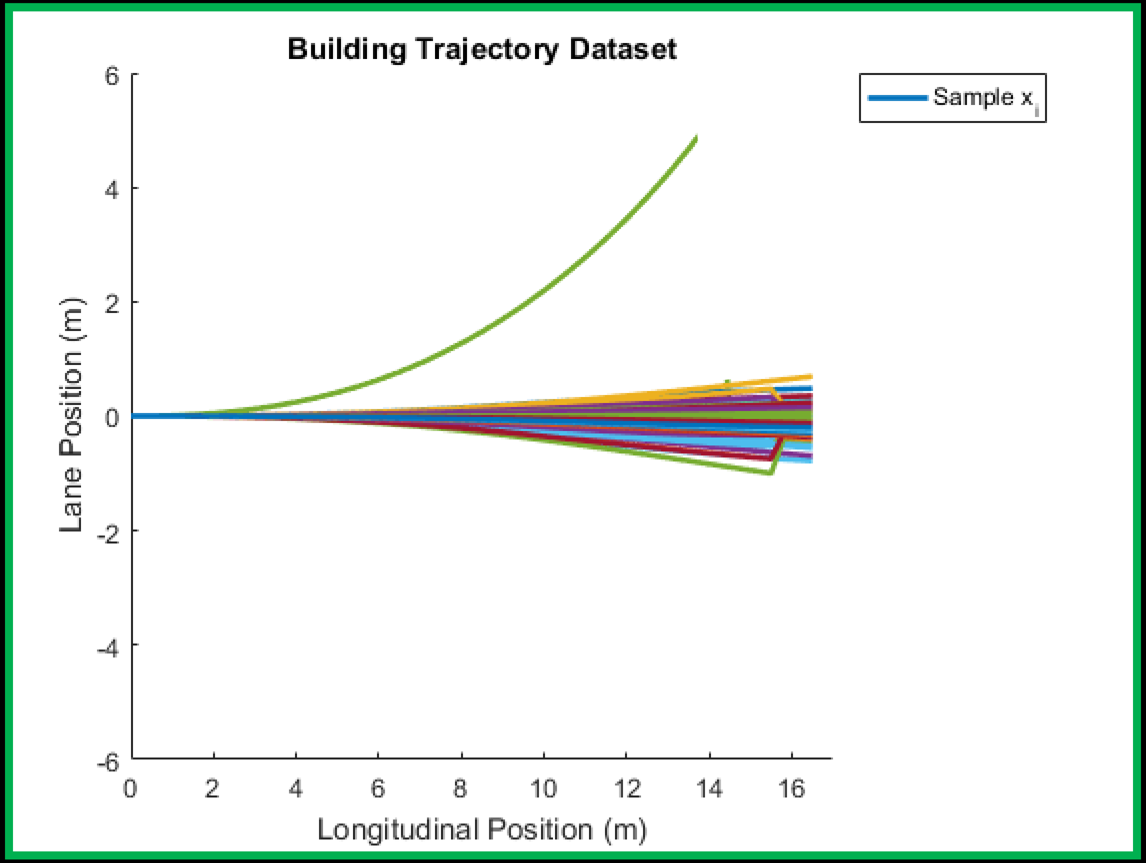
$$\text{argmin}_{\Delta \subset \mathbb{R}^n} \ \lambda(\Delta)$$

$$\text{subject to} \quad \hat{P}_X[\Delta] \geq \alpha$$



K. Driggs-Campbell, et al., *Improved Driver Modeling for Human-in-the-Loop Control*, ICRA 2015.
K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

From Data to Empirical Reachable Sets



$$X = \begin{bmatrix} x_1[0] \cdots x_1[T] \\ \vdots \\ x_N[0] \cdots x_N[T] \end{bmatrix}$$

$$x_i[0] = x_0, \forall i$$

K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

From Data to Empirical Reachable Sets

maximize precision

while maintaining accuracy

$$\begin{aligned} &\operatorname{argmin}_{\Delta \subset \mathbb{R}^n} \; \lambda(\Delta) \\ &\text{subject to} \quad \hat{P}_X[\Delta] \geq \alpha \end{aligned}$$

$$\text{where } \hat{P}_X[\Delta] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\{x_i \in \Delta\}$$

K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

From Data to Empirical Reachable Sets

maximize precision

while maintaining accuracy

$$
\begin{aligned}
\text{argmin}_{\bar{x},\underline{x},b} \quad & \lambda(\bar{x}, \underline{x}) \\
\text{subject to} \quad & b_i(\bar{x} - x_i) \geq 0 \\
& b_i(\underline{x} - x_i) \leq 0 \\
& \textstyle\sum_i b_i \geq N(1-\alpha)
\end{aligned}
$$

where $b_i \in \{0,1\}$

K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

From Data to Empirical Reachable Sets

maximize precision

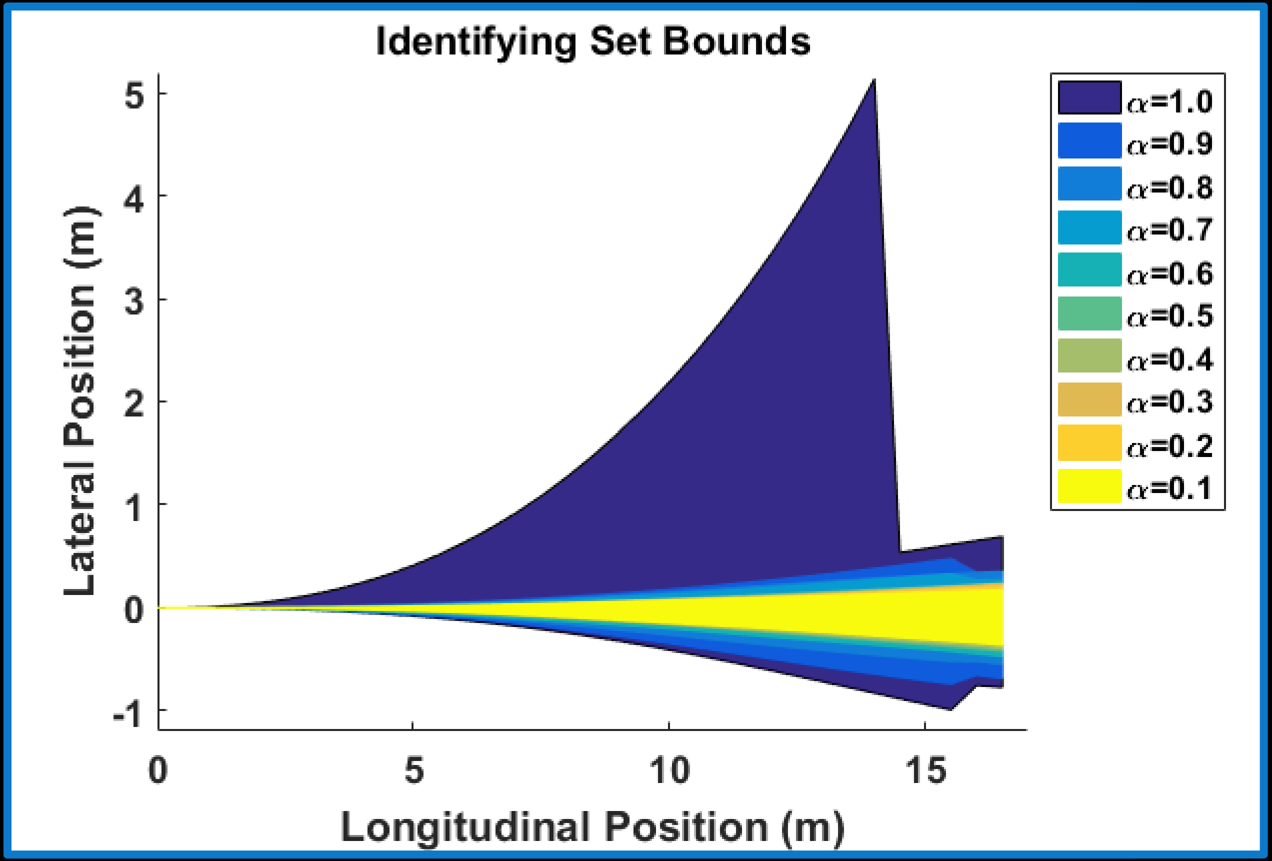
while maintaining accuracy

$$
\begin{aligned}
&\operatorname{argmin}_{\bar{x},\underline{x},b} && \lambda(\bar{x},\underline{x}) \\
&\text{subject to} && \bar{x} - x_i \geq (1 - b_i)(x_{min} - x_i) \\
& && \underline{x} - x_i \leq (1 - b_i)(x_{max} - x_i) \\
& && \textstyle\sum_i b_i \geq N(1-\alpha)
\end{aligned}
$$

where $b_i \in \{0,1\}$

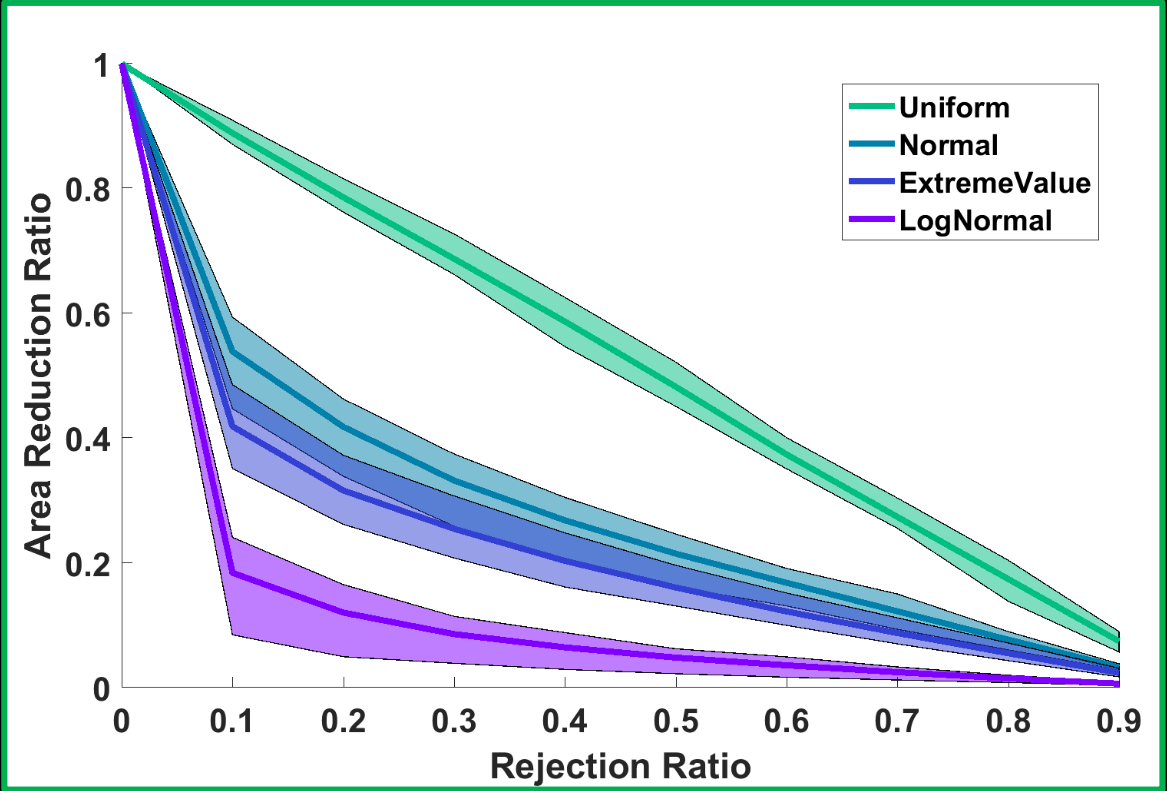
K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

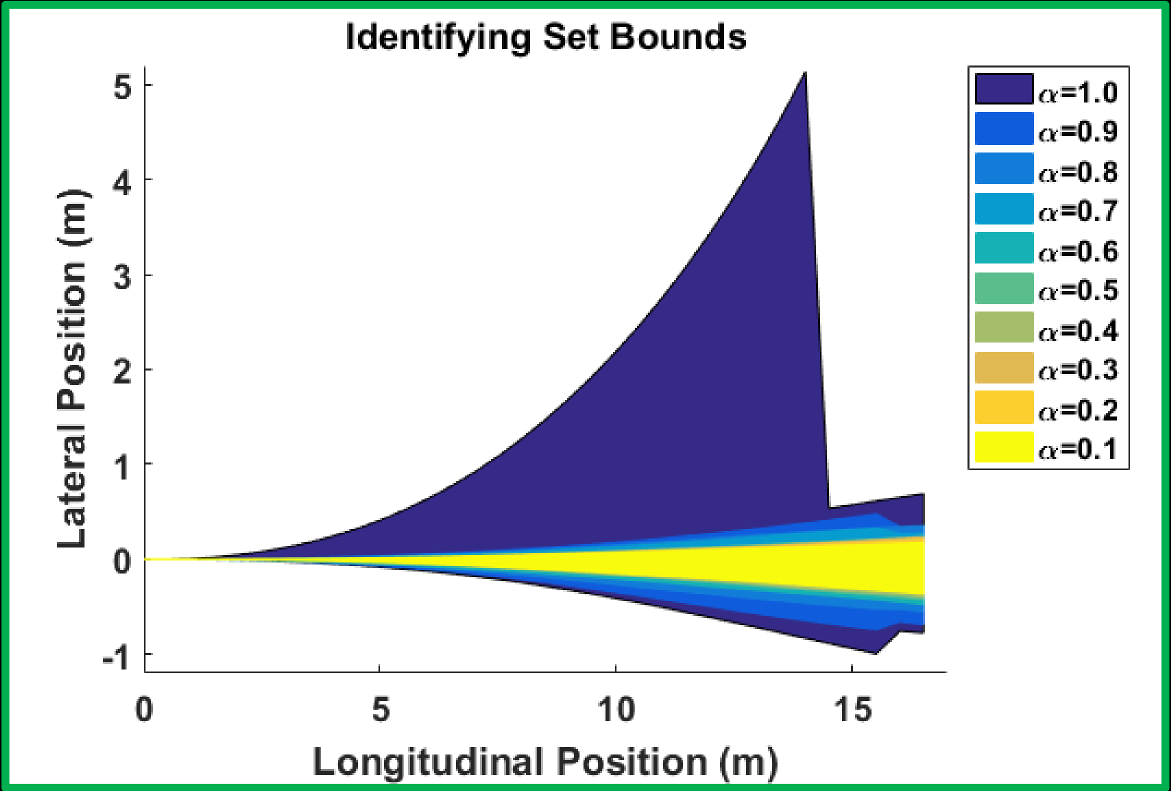
From Data to Empirical Reachable Sets



K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

Solvers, Monotonicity, and Distributions

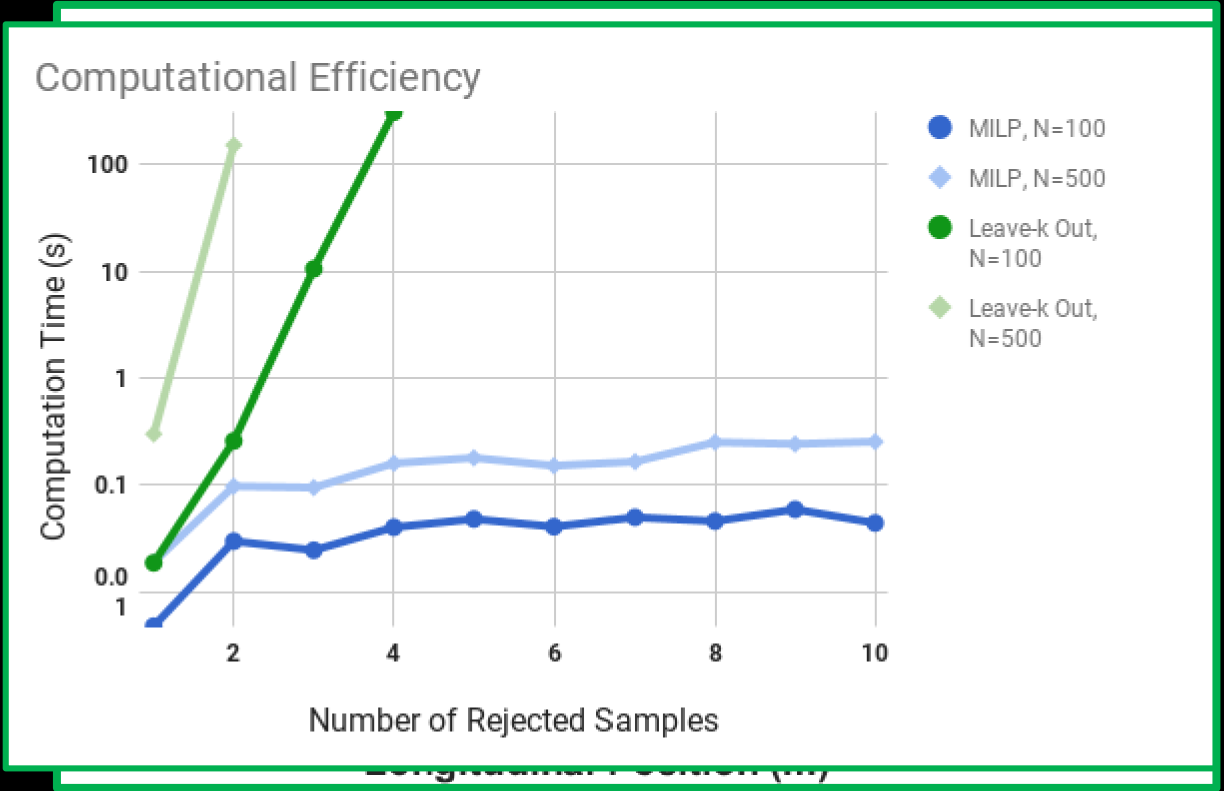




K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

Solvers, Monotonicity, and Distributions

- For n binary variables (samples), we have 2^n optimization problems
- Using branch and bound method, search much fewer combinations
- Rejection ratio increases significant increases in computation time



K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

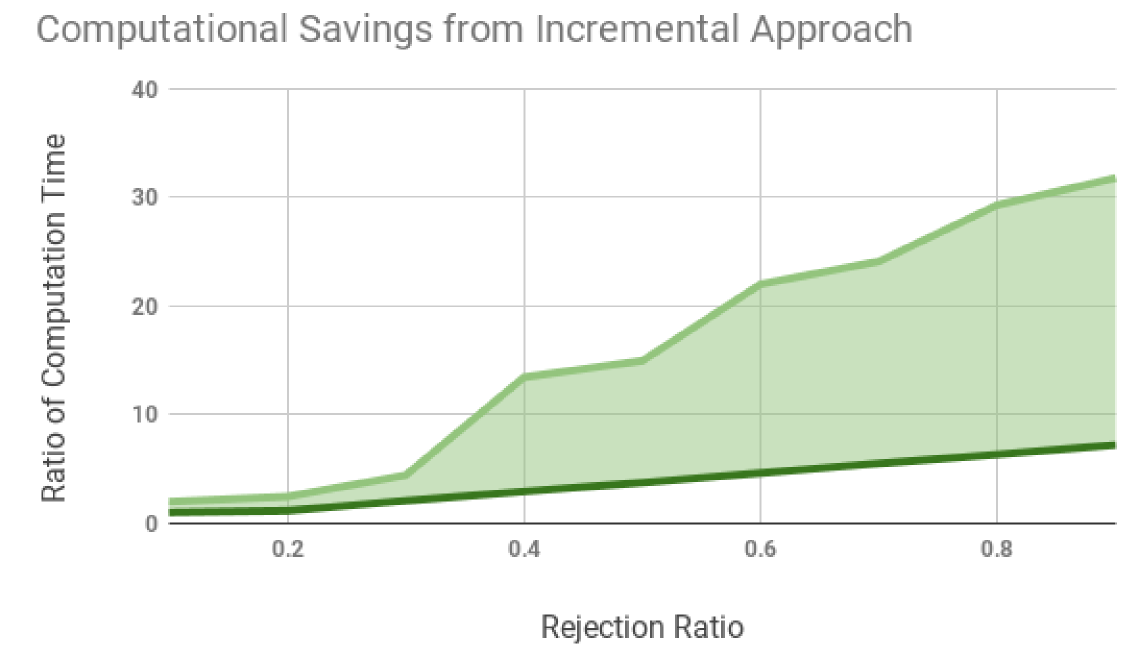
ILLINOIS

Solvers, Monotonicity, and Distributions

If the input data is unimodal, then the output is monotonic, in the sense that:

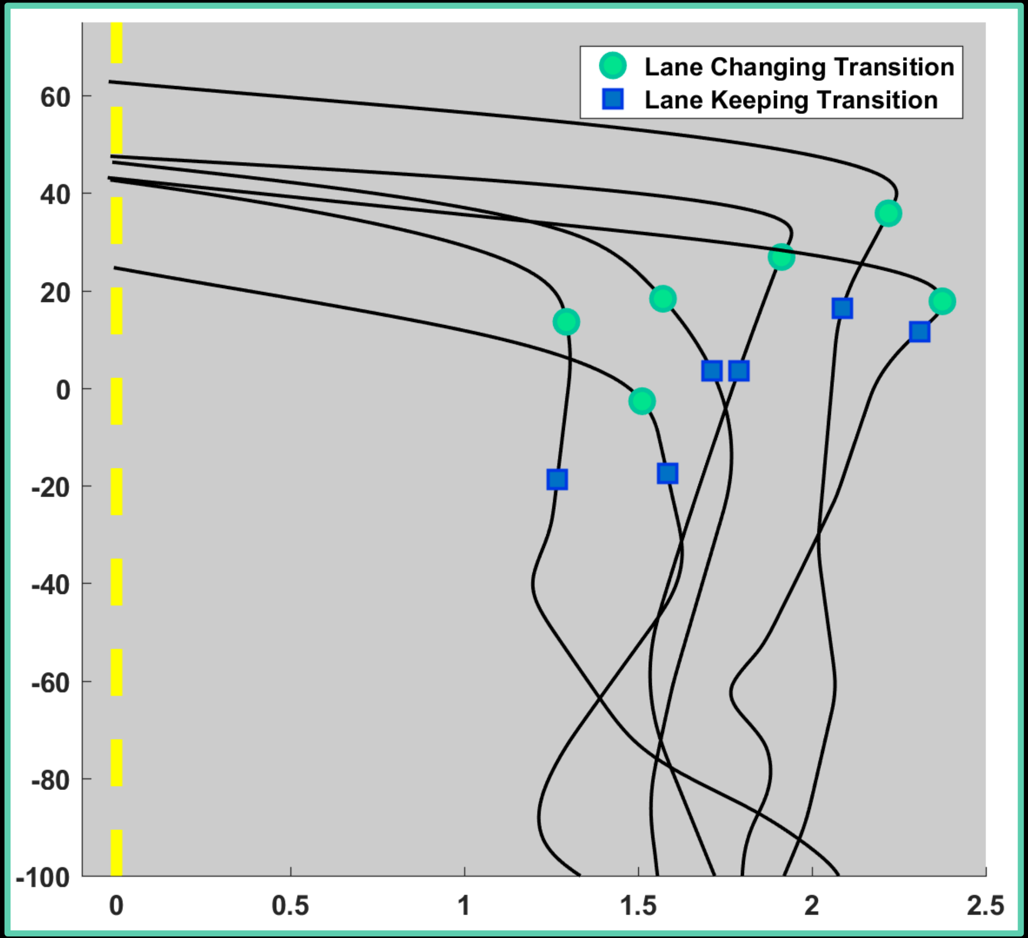
$$\Delta(\alpha_p) \subseteq \Delta(\alpha_q), \forall \alpha_p \leq \alpha_q$$

Assuming unimodality, iteratively remove points until the rejection ratio is met to efficiently find the global optimum



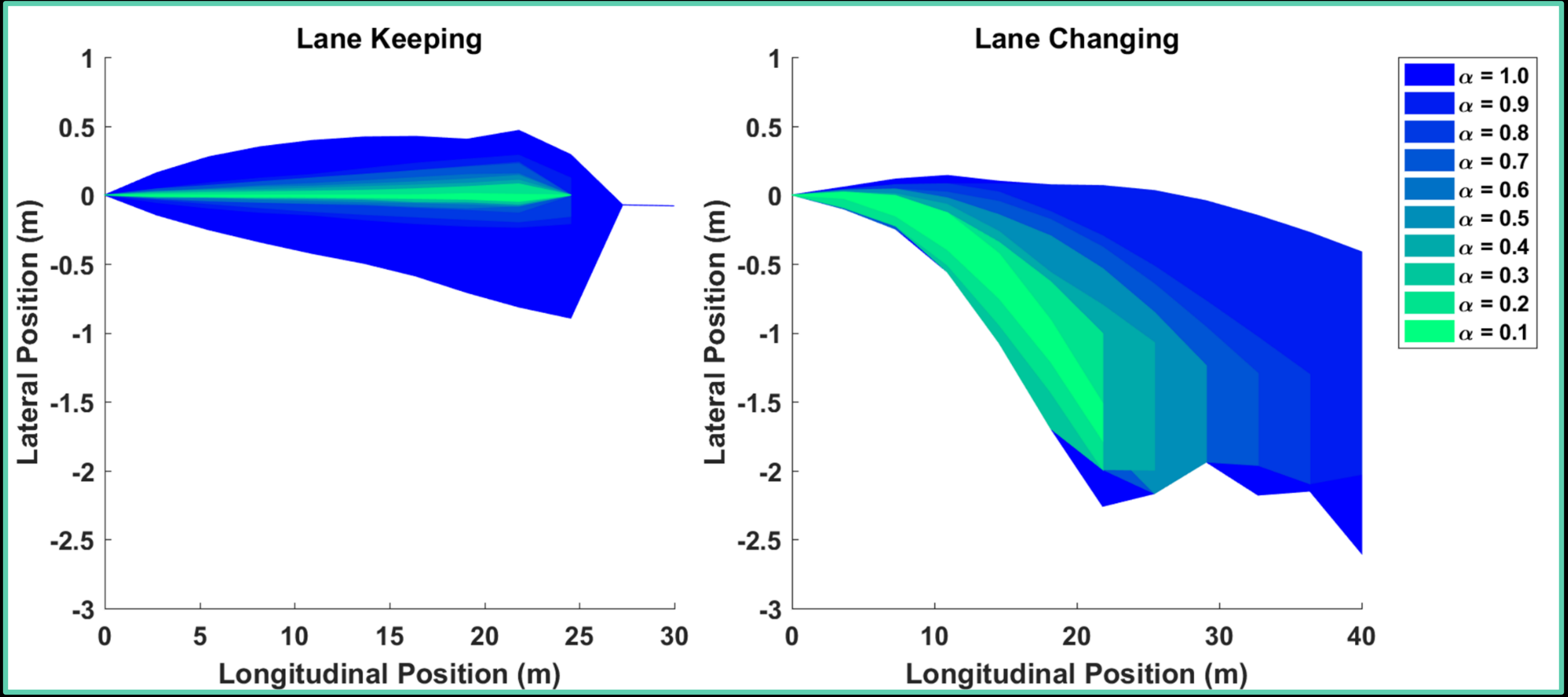
K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

Detecting Driver Modes



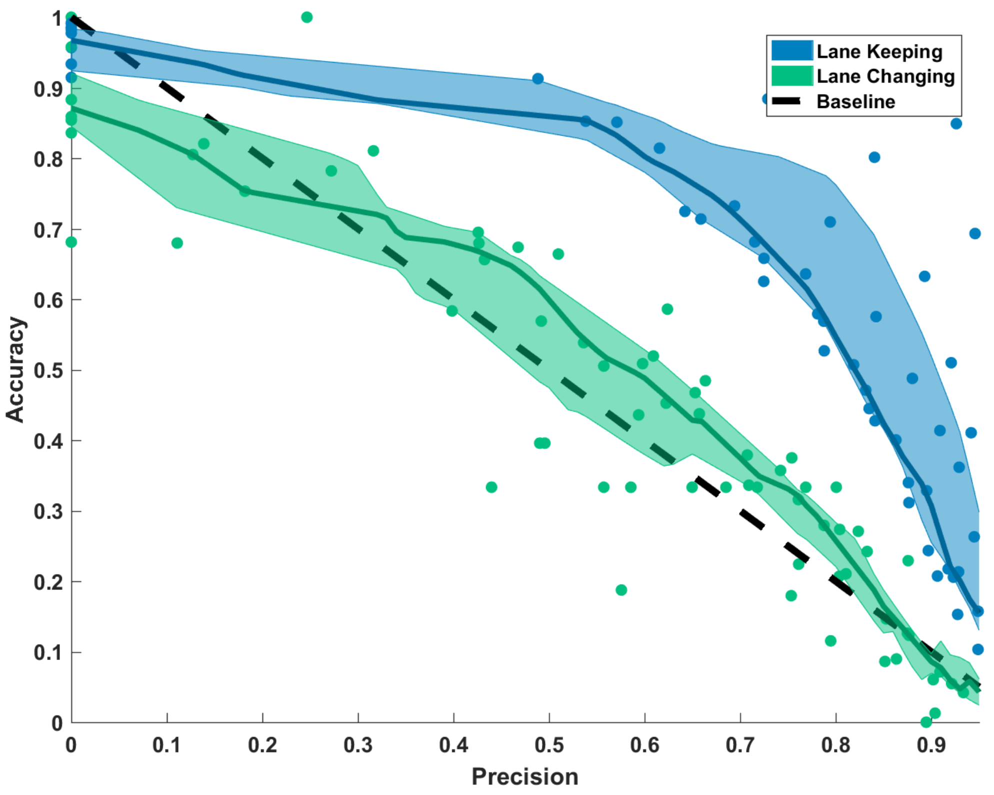
K. Driggs-Campbell, et al., *Identifying Modes of Intent from Driver Behaviors in Dynamic Environments,* ITSC 2015.

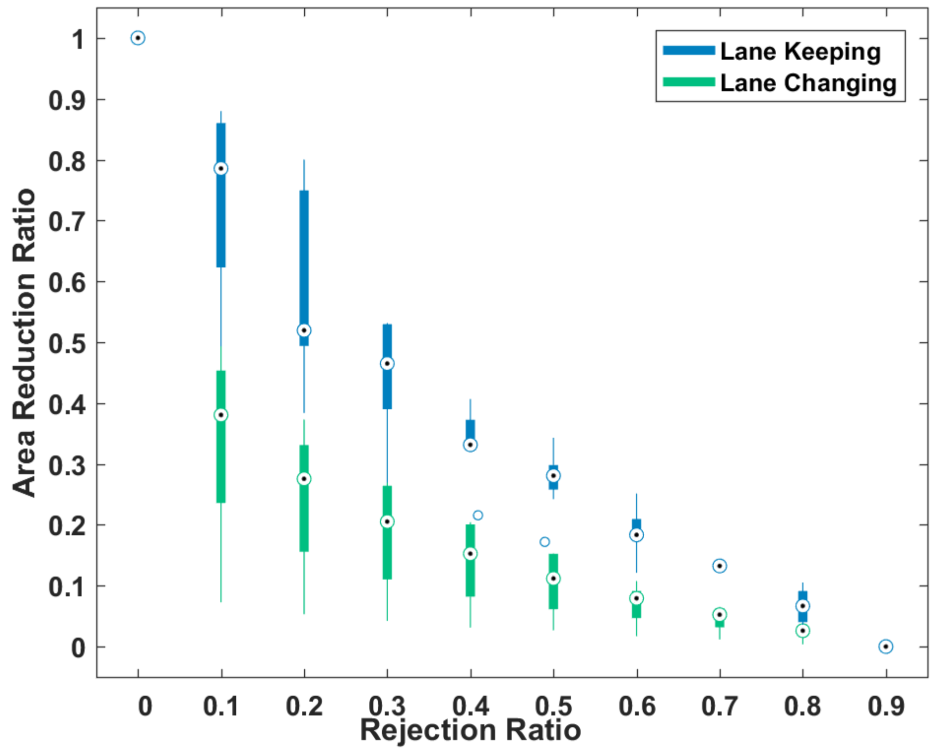
Results on Lane Changing Example



K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

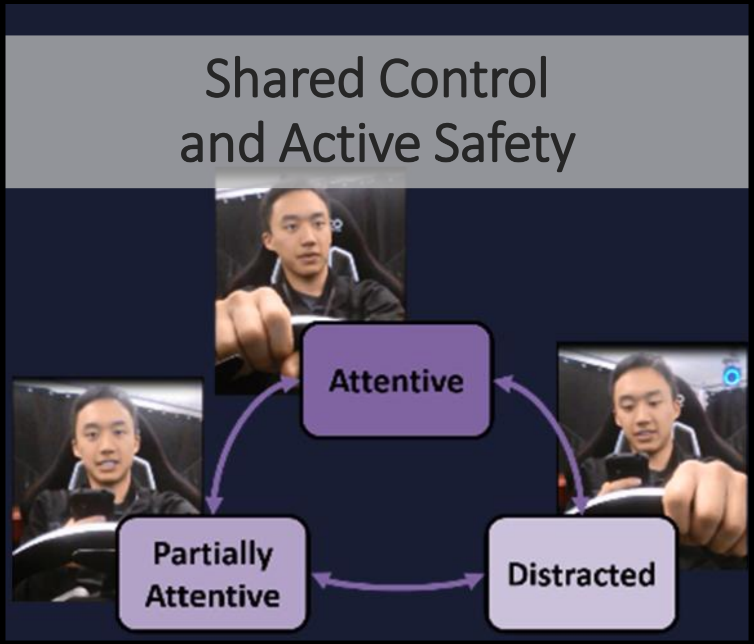
Results on Lane Changing Example

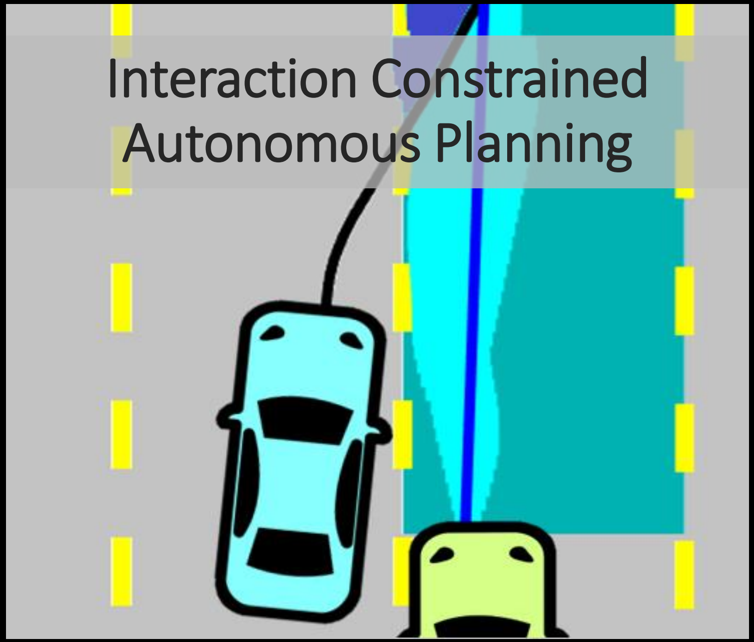


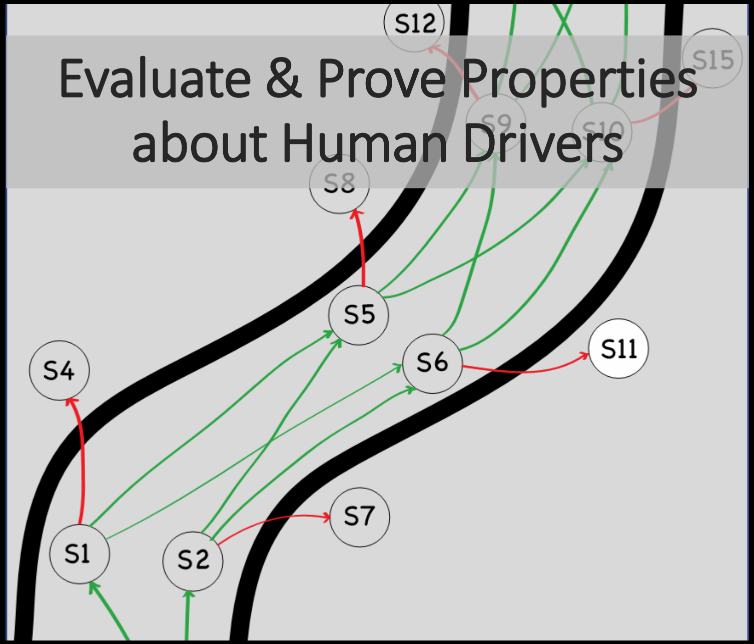


K. Driggs-Campbell, et al., *Robust, Informative Human-in-the-Loop Predictions via Empirical Reachable Sets*, in Transactions on Intelligent Vehicles, 2018.

Applications for ERS



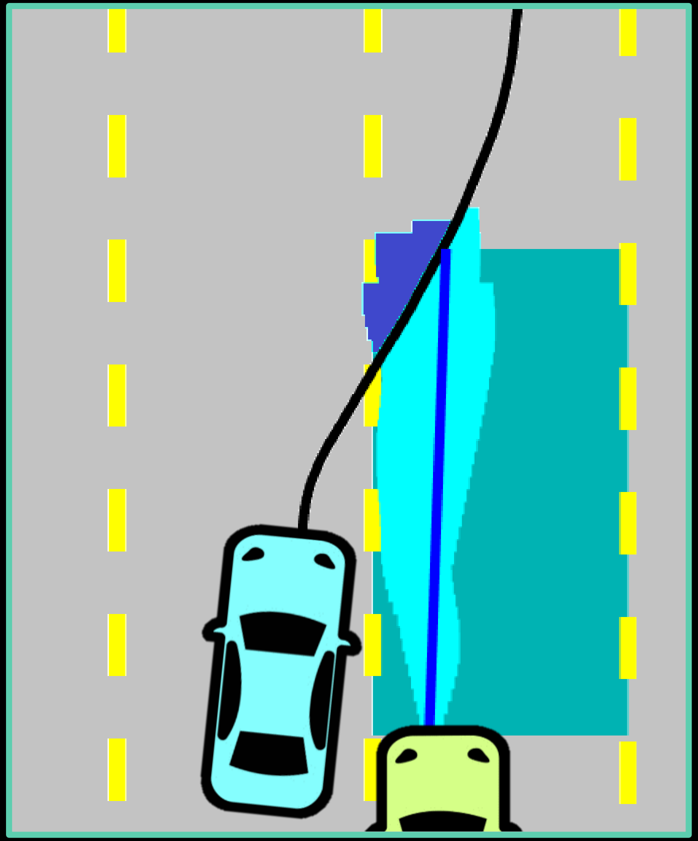




K. Driggs-Campbell, et al., *Improved Driver Modeling for Human-in-the-Loop Control*, ICRA 2015.
V. Shia, Y. Gao, R. Vasudevan, **K. Driggs-Campbell,** et al. *Semi-Autonomous Vehicular Control Using Driver Modeling*, in Transactions on ITS 2014.
K. Driggs-Campbell, et al., *Integrating Intuitive Driver Models in Autonomous Planning for Interactive Maneuvers,* in Transactions on ITS 2017.
D. Sadigh, **K. Driggs-Campbell**, et al. *Data-driven probabilistic modeling and verification of human driver behavior*, in AAAI 2014.

Interaction Constrained Autonomous Planning

If we can predict likely driver responses in cooperative maneuvers, can we design autonomous systems that can effectively integrate with human drivers?



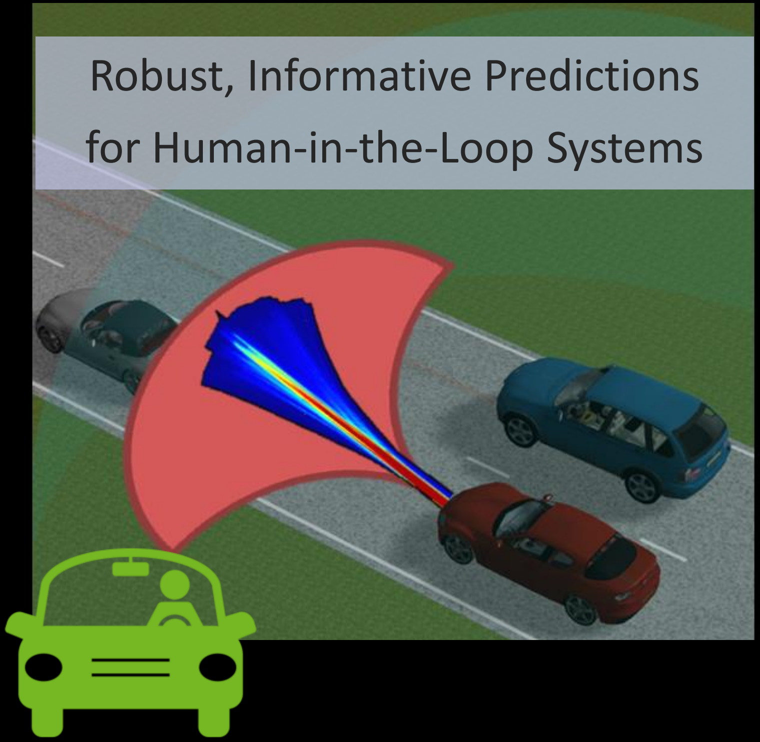
Using robust, predictive models of drivers, we can:

- Effectively predicts drivers' merging responses
- Incorporating these sets as planning constraints results in more human-like motion
- Human-inspired controllers increases predictability by ~40%

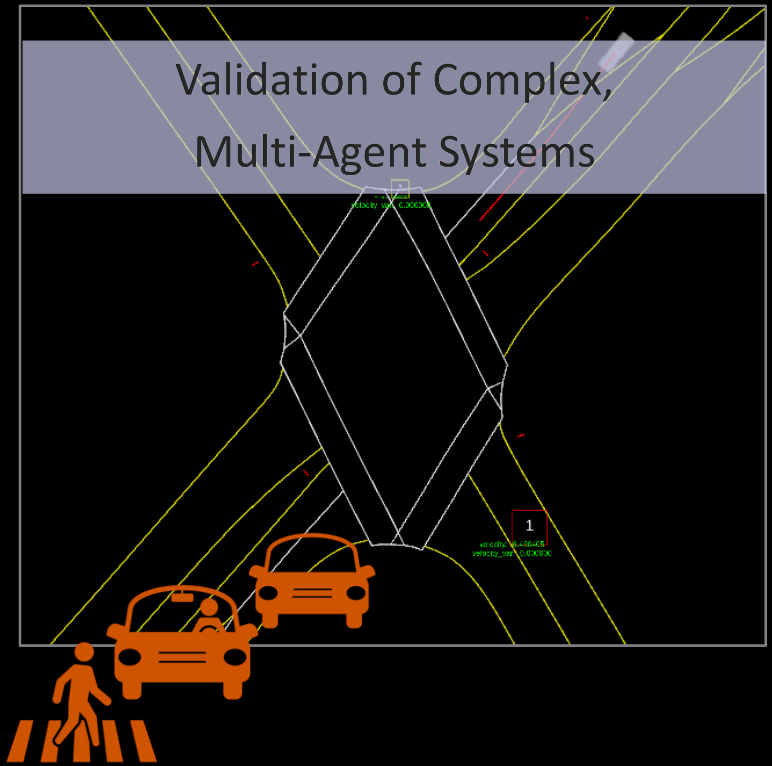
K. Driggs-Campbell, et al., *Integrating Intuitive Driver Models in Autonomous Planning for Interactive Maneuvers,* in Transactions on Intelligent Transportation, 2017.

K. Driggs-Campbell, et al. *Communicating Intent on the Road Through Human-Inspired Control Schemes*, IROS 2016.

Roadmap


Robust, Informative Predictions
for Human-in-the-Loop Systems

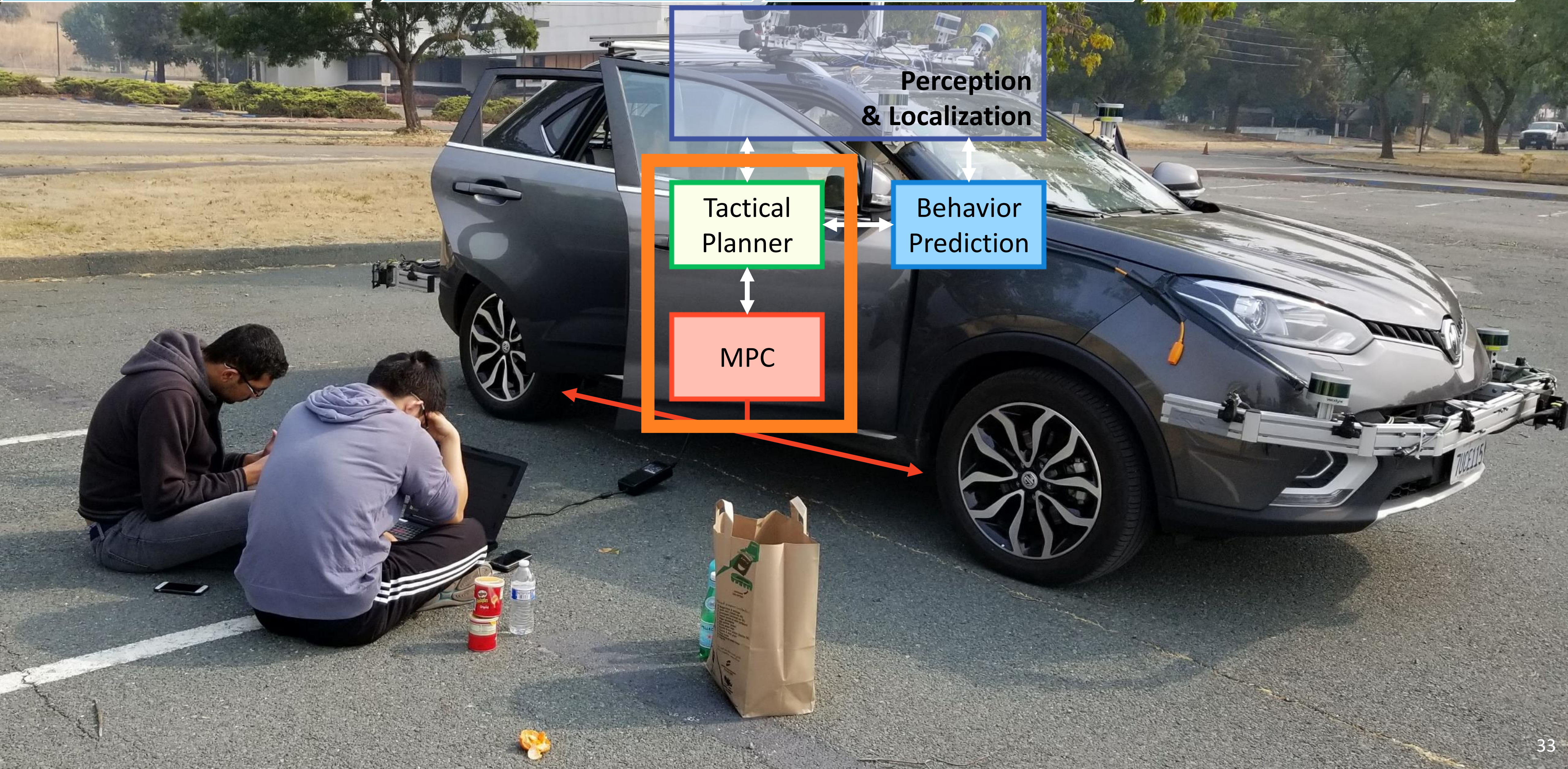

Towards Safe and
Efficient Urban Driving


Validation of Complex,
Multi-Agent Systems

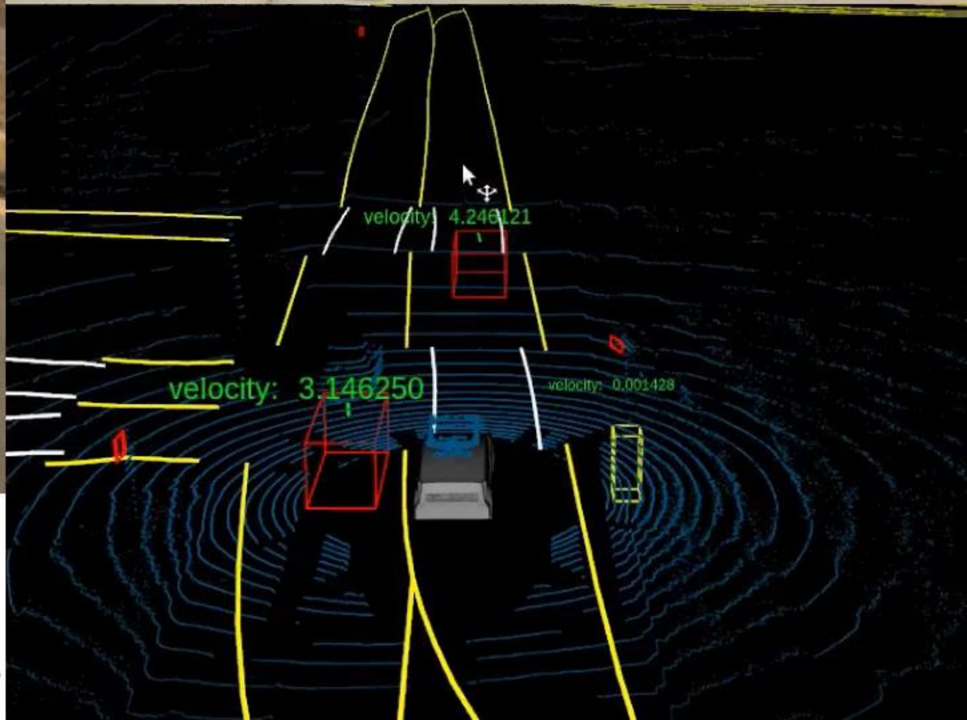

Autonomous
Vehicle (AV)
Testing Zone
GOMENTUM STATION
Redefining Mobility


How do we create a safe and effective autonomous vehicle?

Robust MPC → Safe Imitation Learning → Rigorous Validation → Integration & Implementation



Testing at Gomentum Station

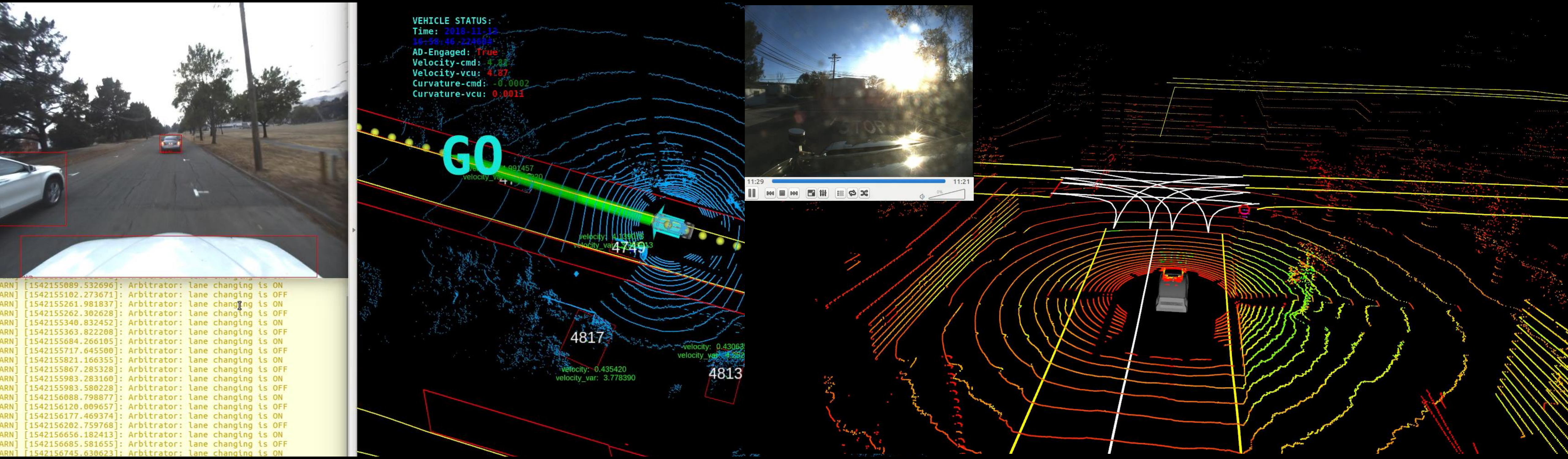


Urban Lane Change Demo at Gomentum Station

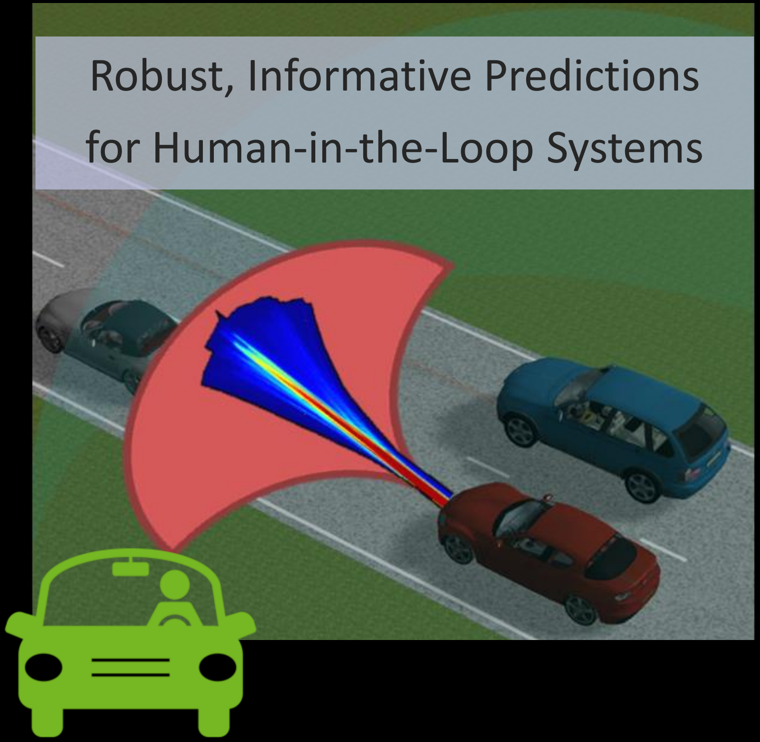
Xiaobai Ma, Michael Kelly,
Katie Driggs-Campbell, and Mykel J. Kochenderfer

SISL
Stanford Intelligent
Systems Laboratory

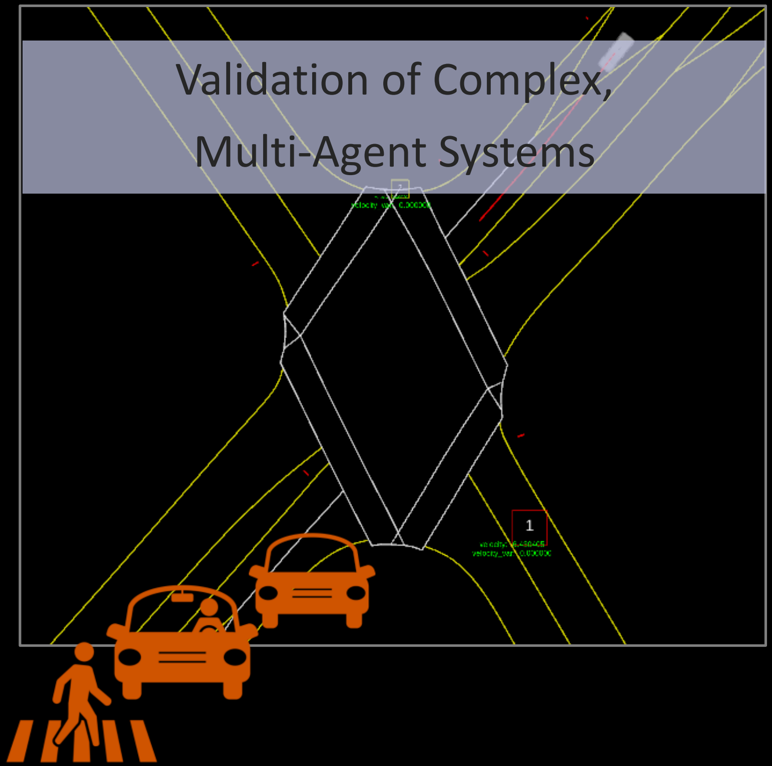
Intervention Scenarios



Roadmap


Robust, Informative Predictions
for Human-in-the-Loop Systems


Towards Safe and
Efficient Urban Driving


Validation of Complex,
Multi-Agent Systems

Low-Probability, High-Risk Events

Hazardous Event Frequencies	
Disengagement Rate	0.12 per 1000 km
Collision Rate	12.5 per 100 million km
Fatality Rate	0.70 per 100 million km

To be meaningful, on the order of billions of kilometers must be driven or simulated.
→ Alternatively, we need an efficient, scalable method for validating complex systems.

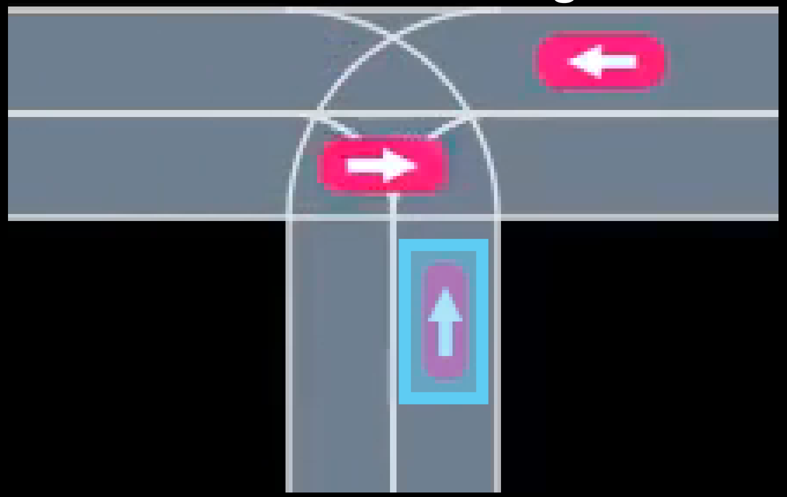
J. Morton, T. Wheeler, and M.J. Kochenderfer. *Closed-Loop Policies for Operational Tests of Safety-Critical Systems.* Under Review 2018.
R. Lee, et al. *Adaptive Stress Testing of Airborne Collision Avoidance Systems*, in DASC 2015.

Low-Probability, High-Risk Events

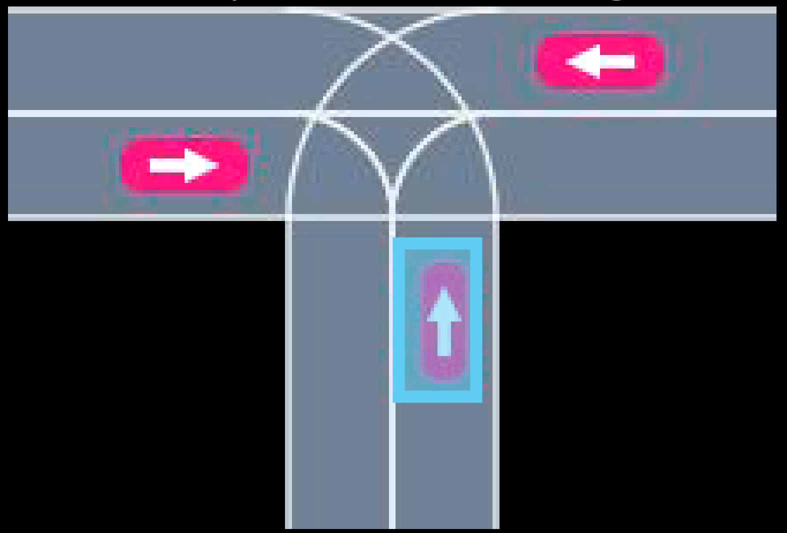
Hazardous Event Frequencies	
Disengagement Rate	0.12 per 1000 km
Collision Rate	12.5 per 100 million km
Fatality Rate	0.70 per 100 million km

To be meaningful, on the order of billions of kilometers must be driven or simulated.
→ Alternatively, we need an efficient, scalable method for validating complex systems.

Traditional Testing

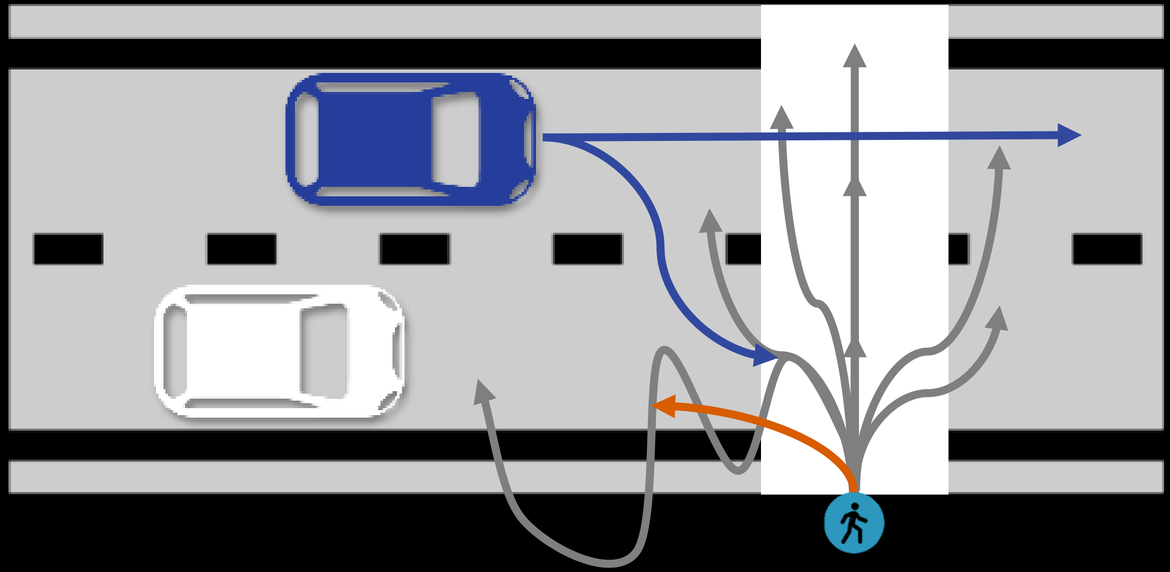


Adaptive Stress Testing



J. Morton, T. Wheeler, and M.J. Kochenderfer. *Closed-Loop Policies for Operational Tests of Safety-Critical Systems.* Under Review 2018.
R. Lee, et al. *Adaptive Stress Testing of Airborne Collision Avoidance Systems*, in DASC 2015.

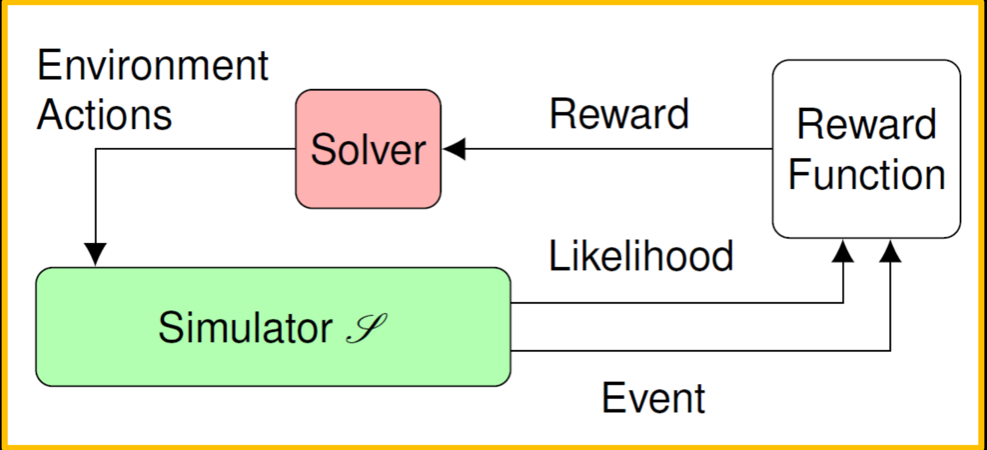
Validation Techniques



- If an agent's motion is discretized, sampling will not give good coverage
- As more agents are added, the number of trajectories to test grows exponentially
- AST aims to identify the most likely failures in your system

AST phrases the validation as a reinforcement learning problem. This provides:

- Sequences of disturbances that will cause the system under test to fail
- Worst case scenarios to qualitatively assess policy performance

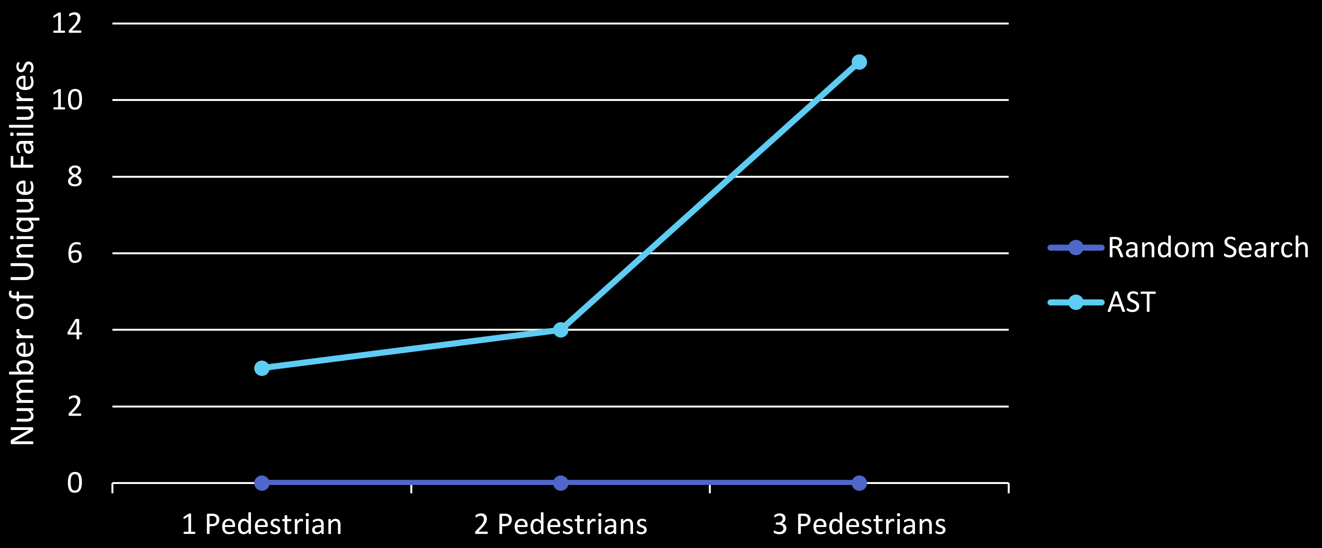


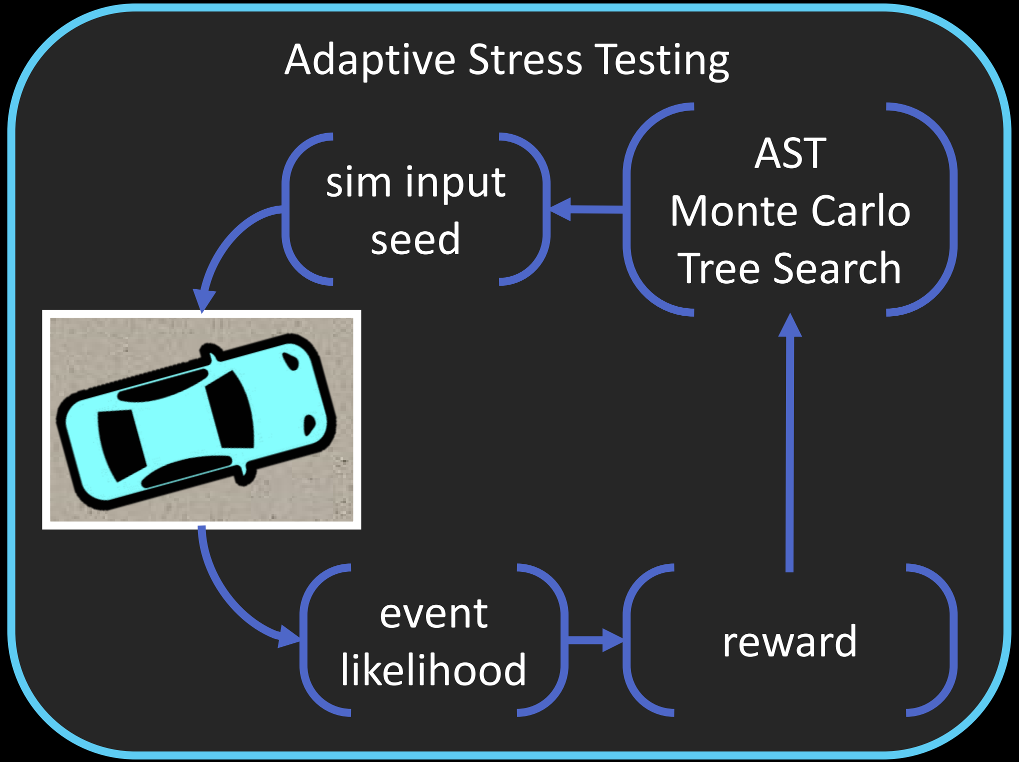
X. Ma,* M. Koren,* A. Corso, **K. Driggs-Campbell**, and M.J. Kochenderfer, *Adaptive Stress Testing Toolbox,* Under Review 2019.

Adaptive Stress Testing

Actively search the system to find the most likely failures in your system:

$$R(s_t, s_{t+1}) = \begin{cases} 0, & \text{if failure} \\ -\infty, & \text{if no failure and } t = T \\ \log P(s_{t+1} | s_t), & \text{if no failure and } t < T \end{cases}$$





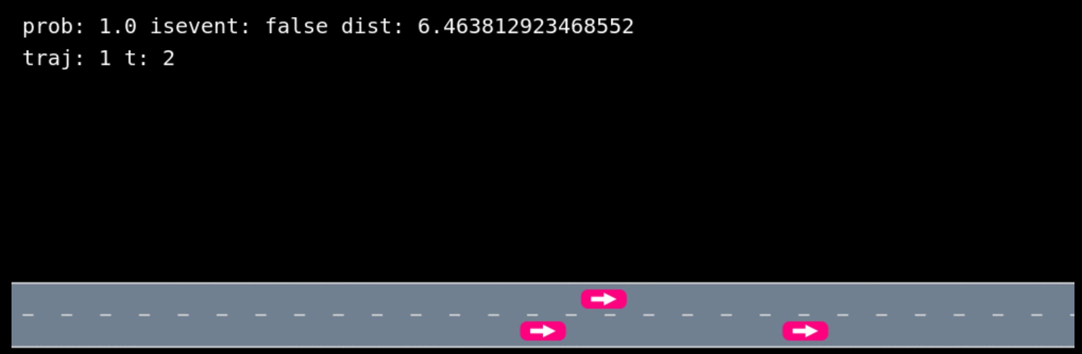
X. Ma,* M. Koren,* A. Corso, **K. Driggs-Campbell**, and M.J. Kochenderfer, *Adaptive Stress Testing Toolbox,* Under Review 2019.

ILLINOIS

AST: Failure Assessment Comparison

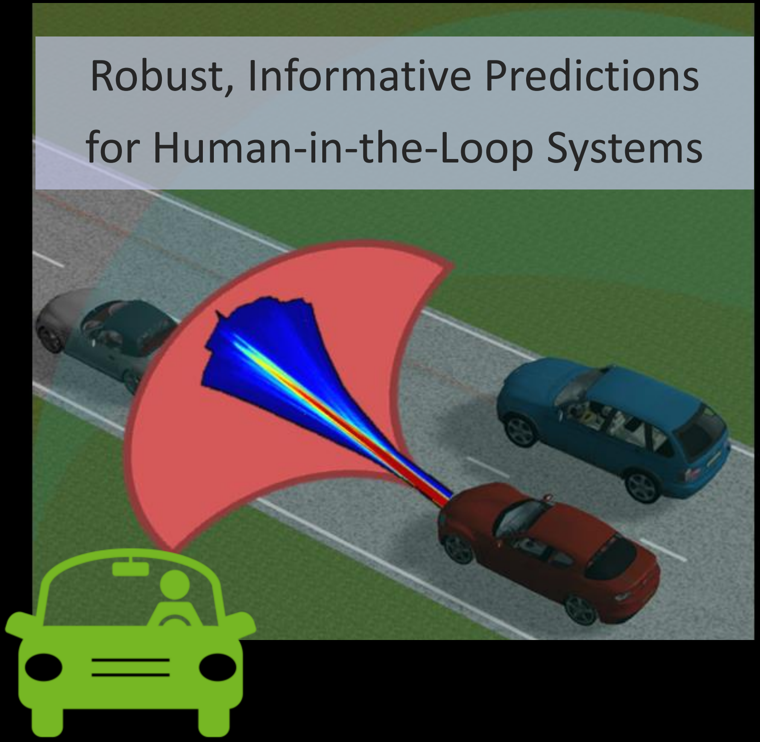
Heuristic / Rule-based System

System with Tracking Errors

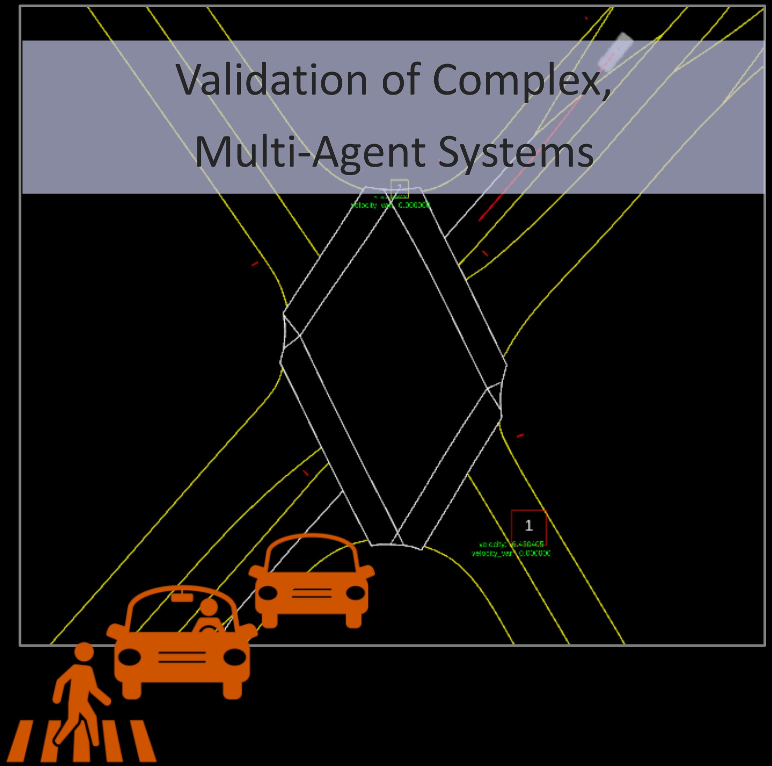


ILLINOIS

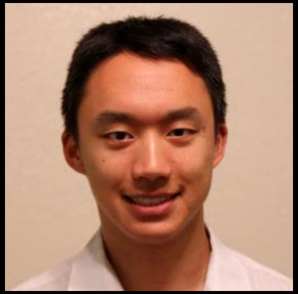
Roadmap


Robust, Informative Predictions
for Human-in-the-Loop Systems


Towards Safe and
Efficient Urban Driving


Validation of Complex,
Multi-Agent Systems
1



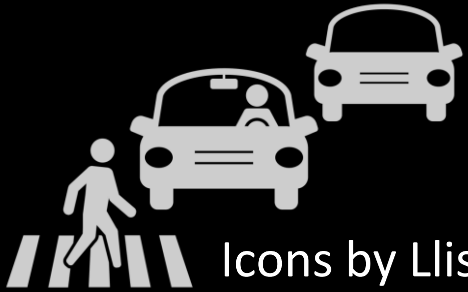




Berkeley
UNIVERSITY OF CALIFORNIA



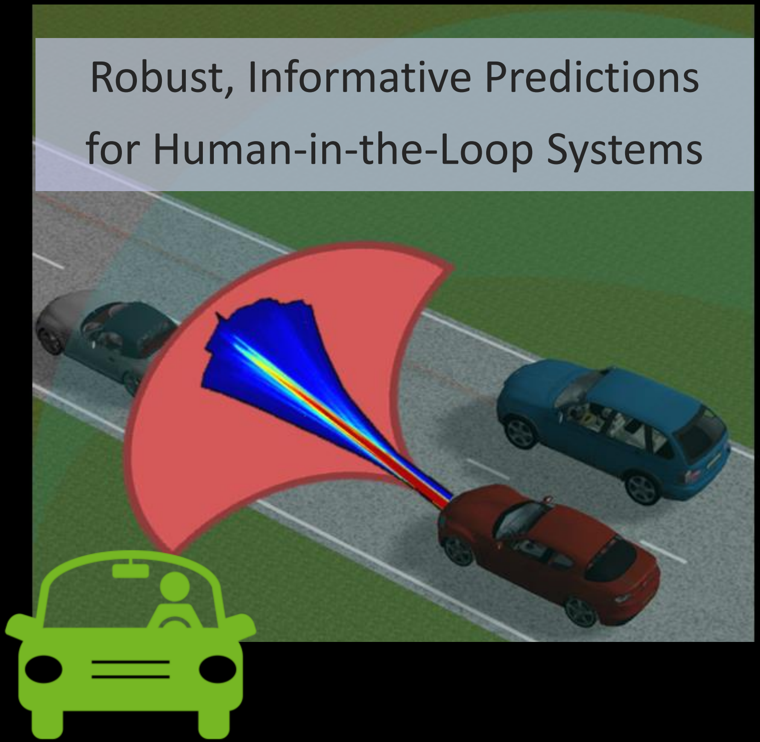


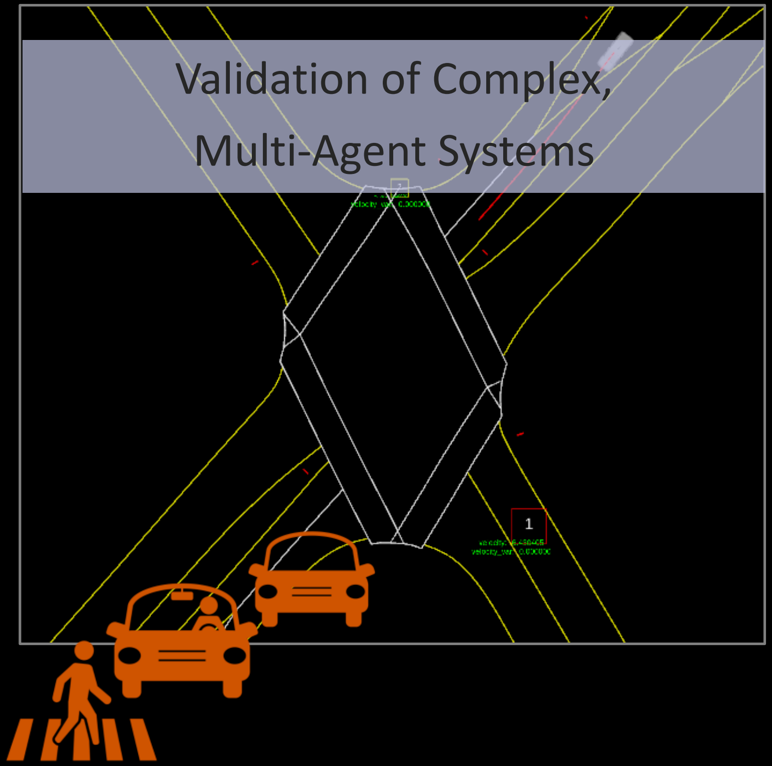

SISL
Stanford Intelligent
Systems Laboratory



Icons by Llisole
on NounProject

Roadmap





thank you!