2019 NSF CYBER-PHYSICAL SYSTEMS PRINCIPAL INVESTIGATORS' MEETING

CPS: Frontier: VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems

S. A. Seshia, R. Bajcsy, B. Hartmann, S. S. Sastry, C. Tomlin (Berkeley),
R. Murray (Caltech), T. Griffiths (Princeton), C. Sturton (UNC Chapel Hill)

VeHICaL http://vehical.org

Caltech

Berkeley
UNIVERSITY OF CALIFORNIA

PRINCETON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Award Numbers: 1545126, 1544714, 1544924

Design of Human Cyber-Physical Systems (h-CPS)

CPS that operate in concert with humans

Semi-Autonomous Driving

UAVs with Human Operators

Robotic Surgery & Medicine

...and other applications.

*Project Goal: To develop a **science of verified co-design** of controllers for semi-autonomous cyber-physical systems and interfaces between humans and cyber-physical components*

Why is this Important?

SAFETY-CRITICAL & MISSION-CRITICAL

Tesla driver dies in first fatal autonomous car crash in US

Hands-off driving faces tough questions

Beck Diefenbach/Reuters

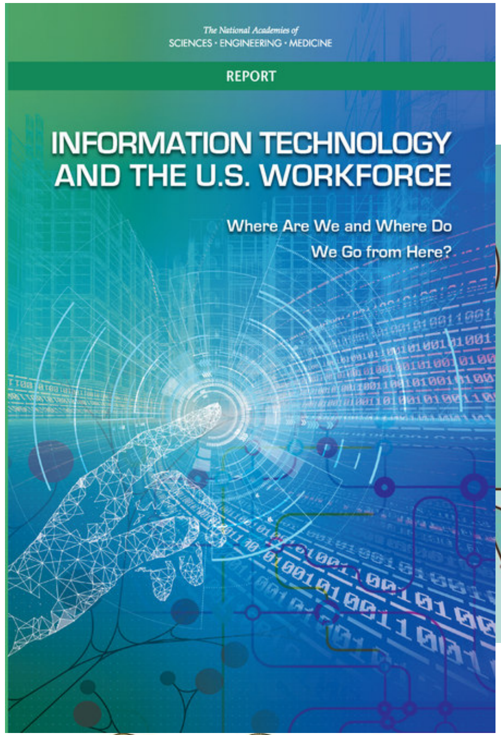
DRIVERLESS CAR SAFETY —

Report: Software bug led to death in Uber's self-driving crash

Sensors detected Elaine Herzberg, but software reportedly decided to ignore her.

TIMOTHY B. LEE - 5/7/2018, 3:12 PM

IMPACT OF AUTOMATION ON WORK/JOBS

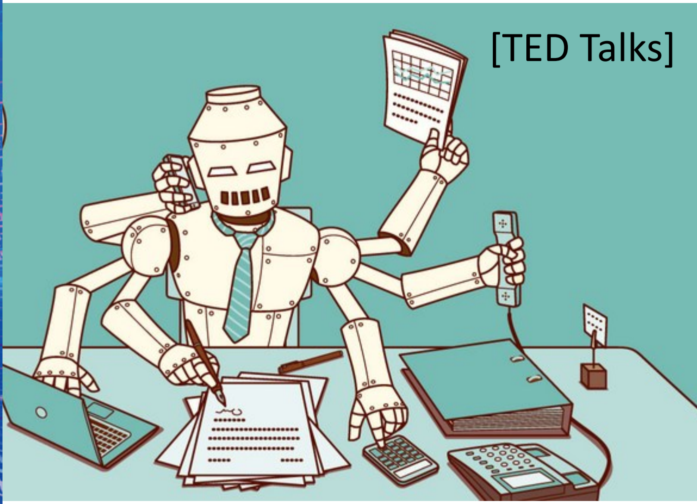


The National Academies of SCIENCES · ENGINEERING · MEDICINE

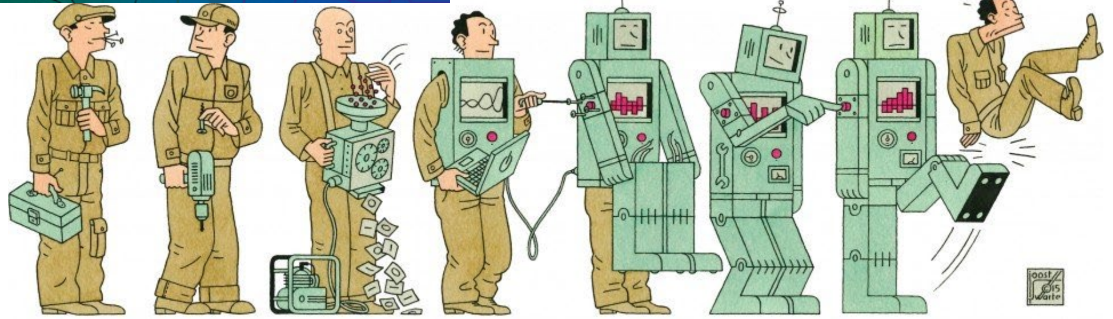
REPORT

INFORMATION TECHNOLOGY AND THE U.S. WORKFORCE

Where Are We and Where Do We Go from Here?

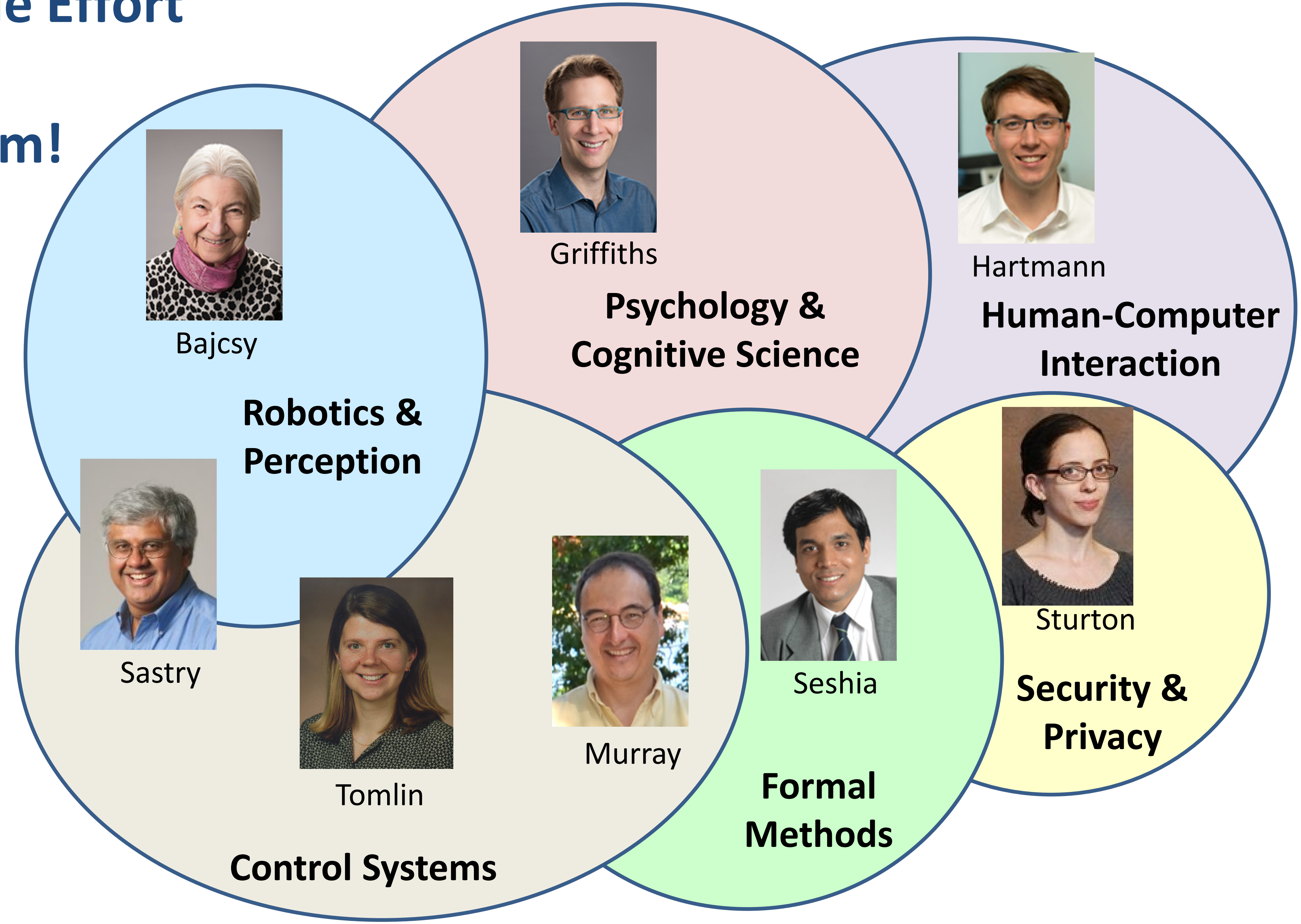


[TED Talks]



[MIT Technology Review]

An Frontier-Scale Effort needs an Inter-Disciplinary Team!



Bajcsy

Robotics & Perception

Griffiths

Psychology & Cognitive Science

Hartmann

Human-Computer Interaction

Sastry

Tomlin

Murray

Control Systems

Seshia

Formal Methods

Sturton

Security & Privacy

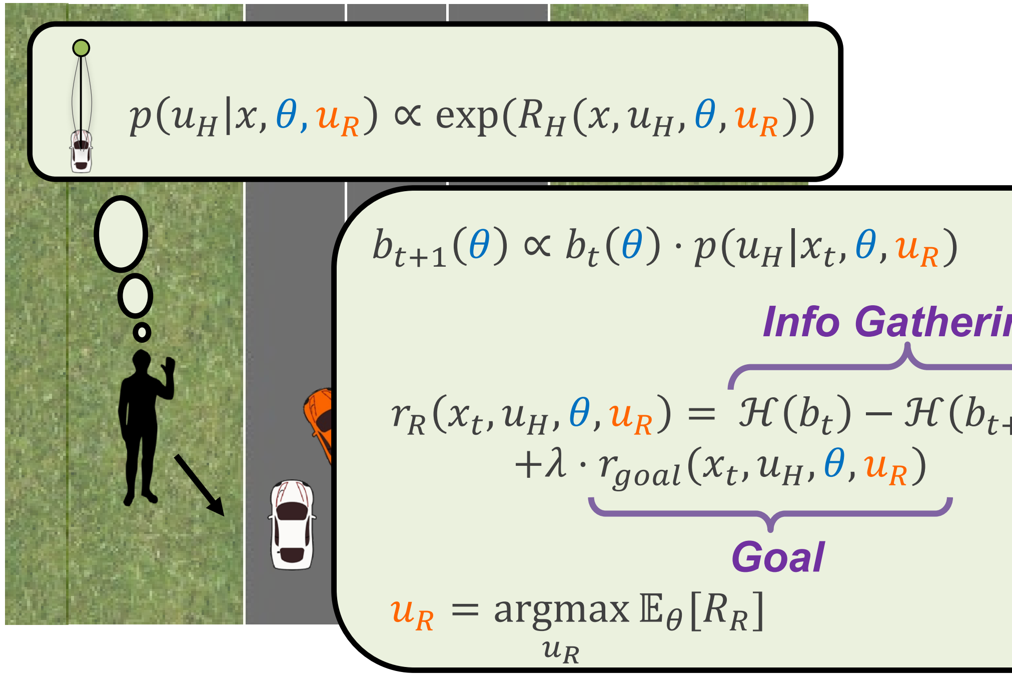
Key Envisioned Contributions to CPS Science

- Developing a Science of Co-Design of Human Interfaces and Control
  - Turning design of h-CPS from an art to a science by systematic design and verification of human interfaces
- Making Uncertainty a first-class citizen in Verification and Control
  - New algorithms and models to deal with uncertainty in CPS dynamics and CPS design
- Bridging the Schism between Model-Based Design and Data-Driven Methods
  - A new design methodology for CPS that blends data-driven learning with formal modeling and proof engines

Design for
Effective Communication between Humans and Automation

Interaction-Aware Control

Leverage human *responses* to estimate human internal state, and learn human model.



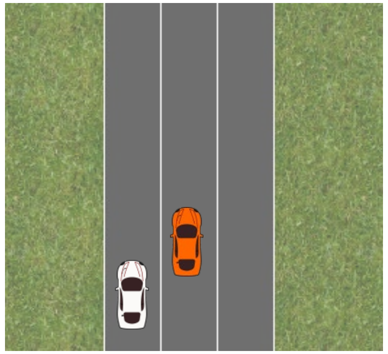
$$p(u_H | x, \theta, u_R) \propto \exp(R_H(x, u_H, \theta, u_R))$$

$$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_H | x_t, \theta, u_R)$$

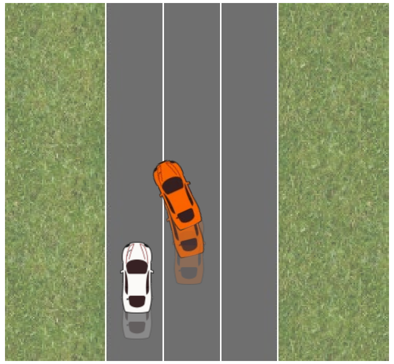
$$r_R(x_t, u_H, \theta, u_R) = \overbrace{\mathcal{H}(b_t) - \mathcal{H}(b_{t+1})}^{\textit{Info Gathering}} \\ \underbrace{+\lambda \cdot r_{goal}(x_t, u_H, \theta, u_R)}_{\textit{Goal}}$$

$$u_R = \underset{u_R}{\operatorname{argmax}} \, \mathbb{E}_\theta [R_R]$$

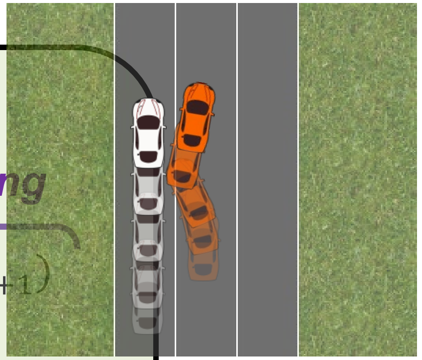
SAMPLE RESULT



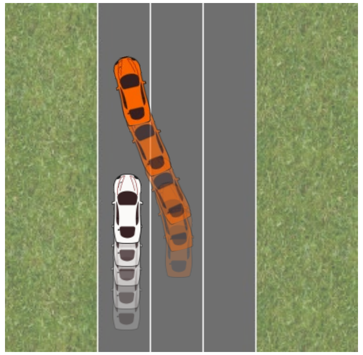
Lane Change



Nudging In

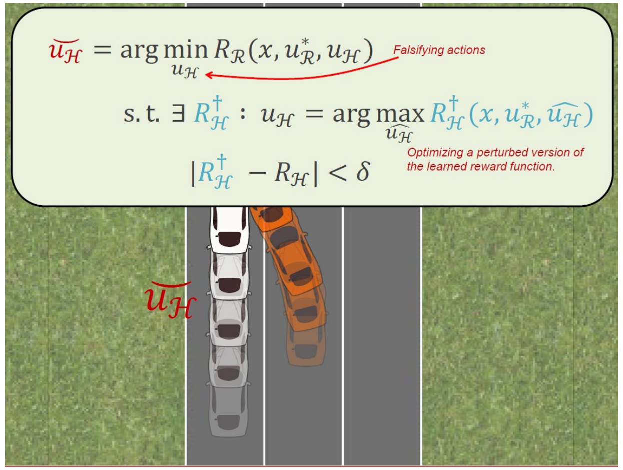


Distracted Human



Attentive Human

VERIFYING ROBUSTNESS



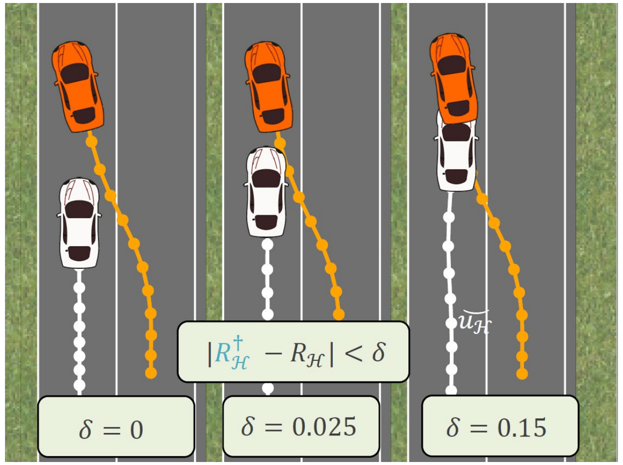
$$\widetilde{u_{\mathcal{H}}} = \arg\min_{u_{\mathcal{H}}} R_{\mathcal{R}}(x, u_{\mathcal{R}}^*, u_{\mathcal{H}})$$

Falsifying actions

$$\text{s.t. } \exists \, R_{\mathcal{H}}^{\dagger} : \; u_{\mathcal{H}} = \arg\max_{\widehat{u_{\mathcal{H}}}} R_{\mathcal{H}}^{\dagger}(x, u_{\mathcal{R}}^*, \widehat{u_{\mathcal{H}}})$$

$$|R_{\mathcal{H}}^{\dagger} - R_{\mathcal{H}}| < \delta$$

Optimizing a perturbed version of the learned reward function.



[Sadigh, Sastry, Seshia, Dragan: RSS, IROS '16]

[Sadigh, Sastry, Seshia: CPHS '18]

Learning and Teaching (Multiple) Task Specifications

Good Communication is Crucial

Demonstrations,
Natural Language
…

Cost Functions,
Logical Specs.
…

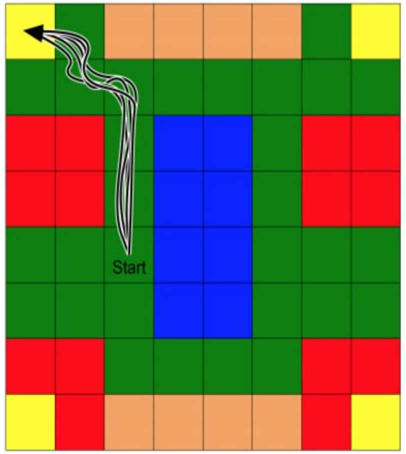
How can we hand off control reliably and intuitively?

Learning Boolean Specifications from Demonstrations

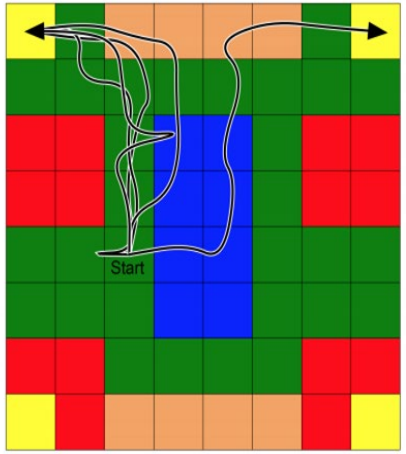
Target Specification:

Go to a yellow tile without going on a red tile. If a blue tile is steped on, step on a brown tile before stepping on a yellow tile

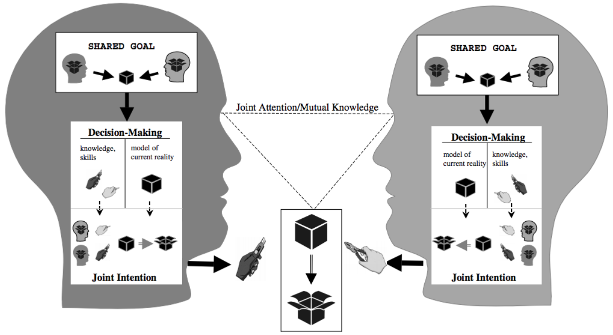
Doing Task



Communicating Task



[Vazquez-Chanlatte, Jha, Ho, et al., NeurIPS'18; CPHS'18]

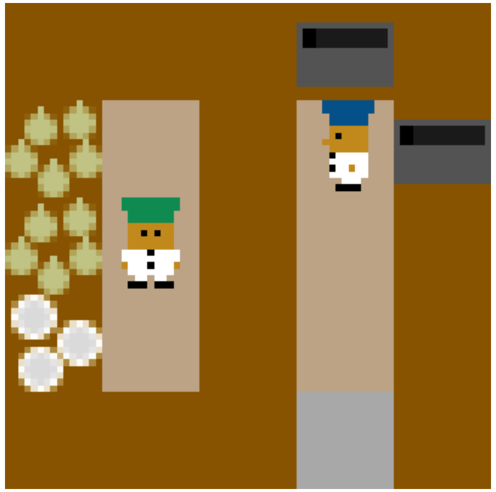


Boolean (logic) specifications:

- Composable
- Non-Markovian tasks
- Leverage formal methods

Humans and machines must coordinate actions and processing

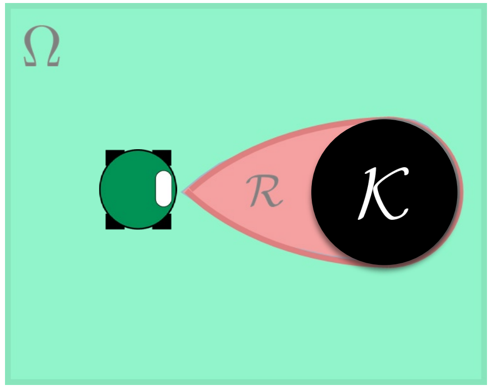
On the Utility of Learning about Humans for Human-AI Coordination



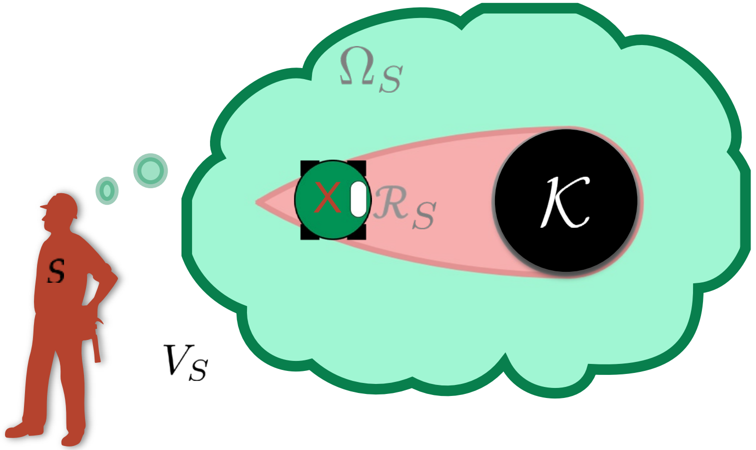
[Carroll, Shah, Ho, et al., NeurIPS'19]

Inferring Supervisor Safe Sets for Human-Robot Teams

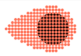
Standard Reachability Safe Set:

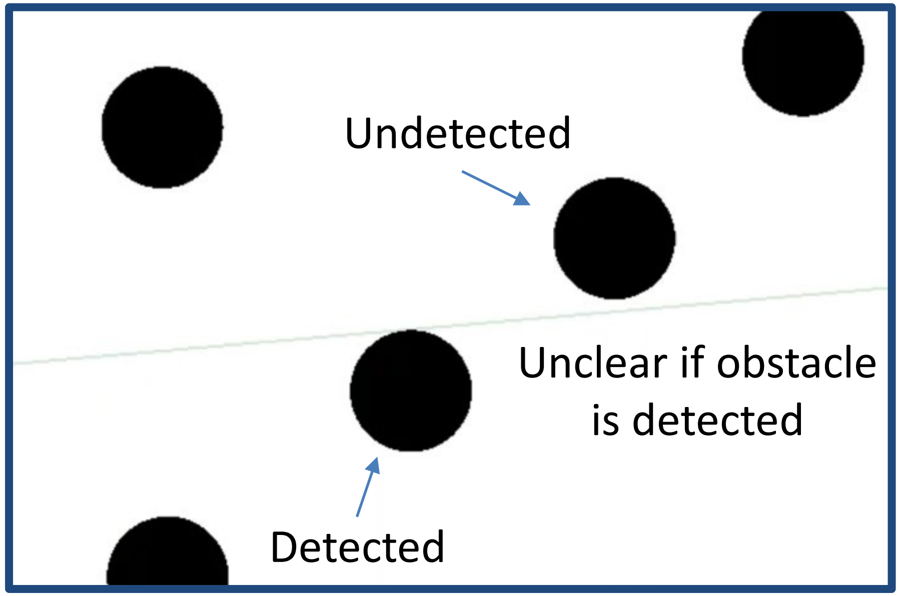


Human's Perceived Safe Set:



- Reachability analysis can compute safe sets
- Humans may perceive safety differently (from each other and from the system designer)
- Our technique can learn these individual safety preferences
- Avoid obstacles using learned safe sets to clearly communicate obstacle detection
- False positives (human thinks obstacle is undetected) are reduced

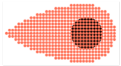
Avoidance with Standard: 

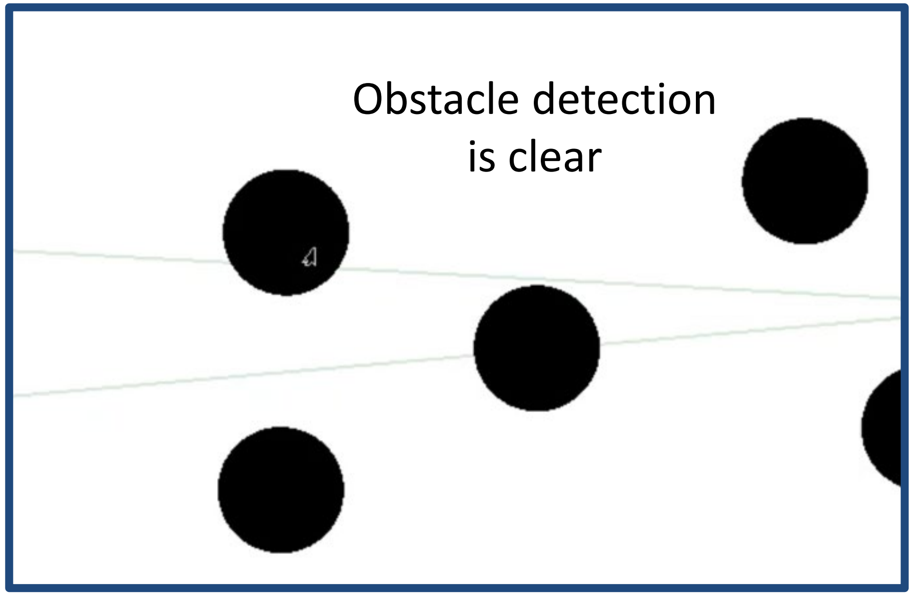


Undetected

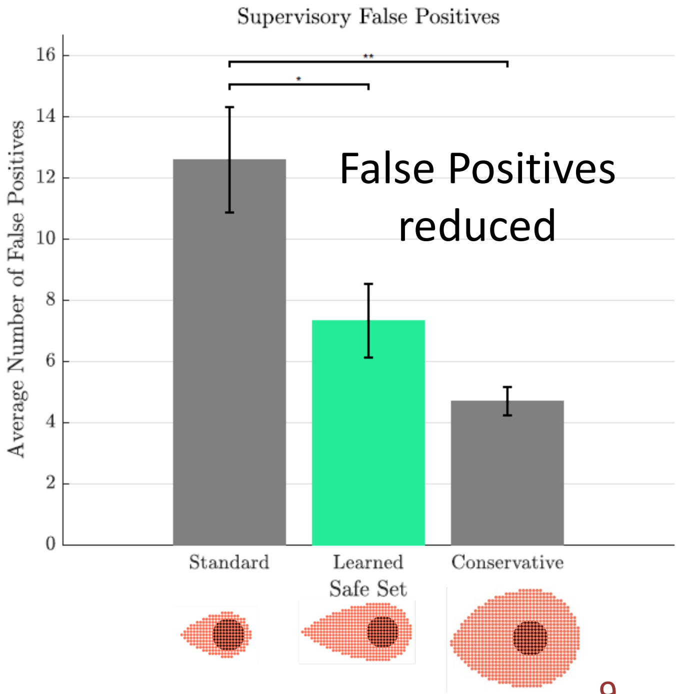
Unclear if obstacle is detected

Detected

Avoidance with Learned: 



Obstacle detection is clear



False Positives reduced

D. McPherson, D. Scobee, J. Menke, A. Yang and S. Sastry,
"Modeling Supervisor Safe Sets for Improving Collaboration in Human-Robot Teams." IROS 2018

Probabilistically Safe Motion Planning Around People

Use Human Models not as Ground Truth, but to Inform *Confidence in Predictions*

Prediction:

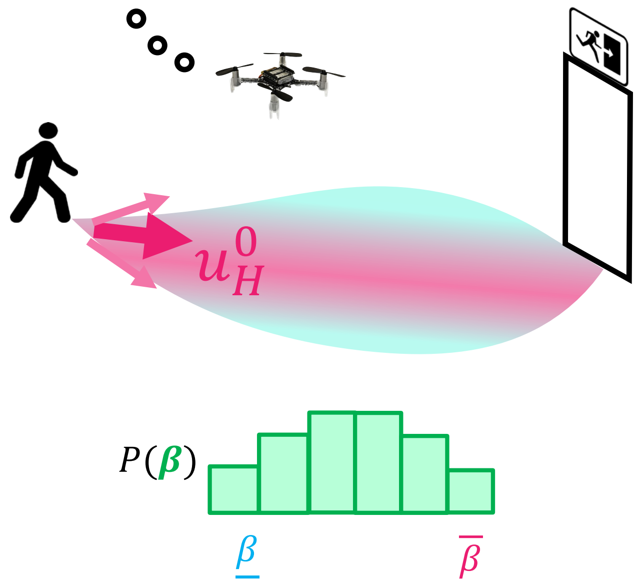
Confidence-aware Human Prediction w/ Boltzmann Model

+

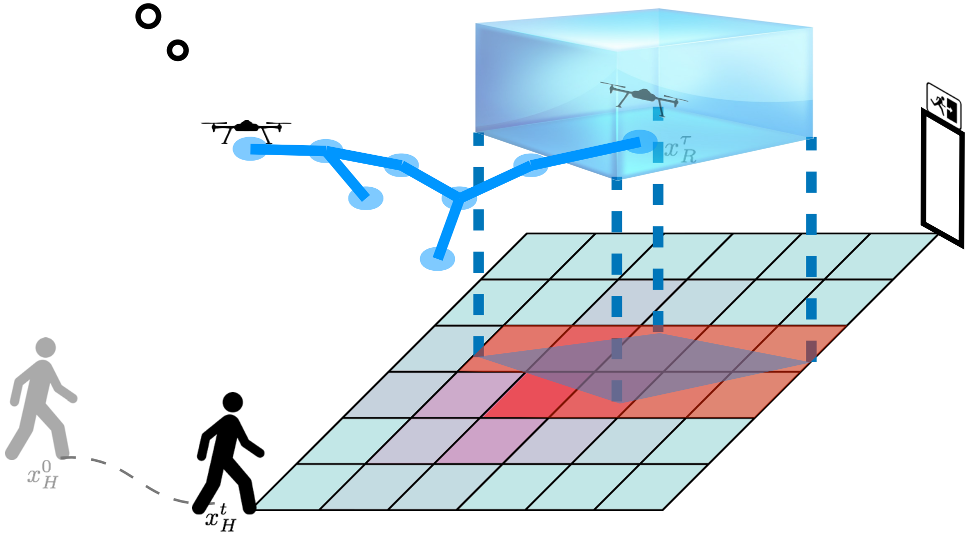
Planning & Control:

Fast and Safe Tracking (FaSTrack)

$$P(u_H^0 \mid x_H^0; \theta, \boldsymbol{\beta}) \propto e^{\boldsymbol{\beta} Q(x_H, u_H; \theta)}$$

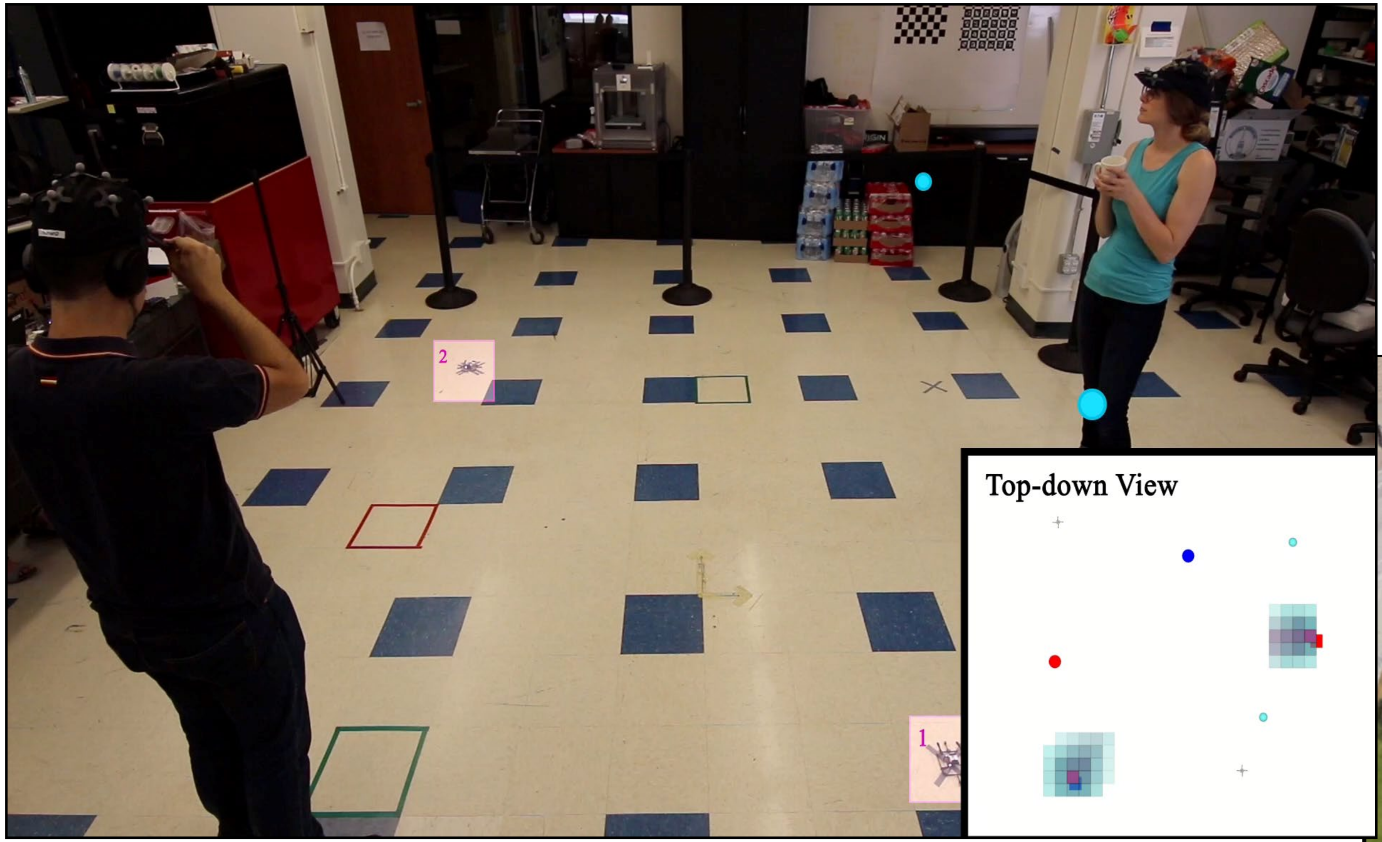


$$P(\mathrm{Crash}(x_R^\tau)) = \mathbb{E}_{\beta,\theta} \int_{\mathcal{H}_\mathcal{E}(x_R^\tau)} dP(x_H^\tau | x_H^t; \beta, \theta)$$



[Andrea Bajcsy, Sylvia Herbert, David Fridovich-Keil, Jaime Fisac, Claire Tomlin 2018]

Probabilistically Safe Motion Planning Around People



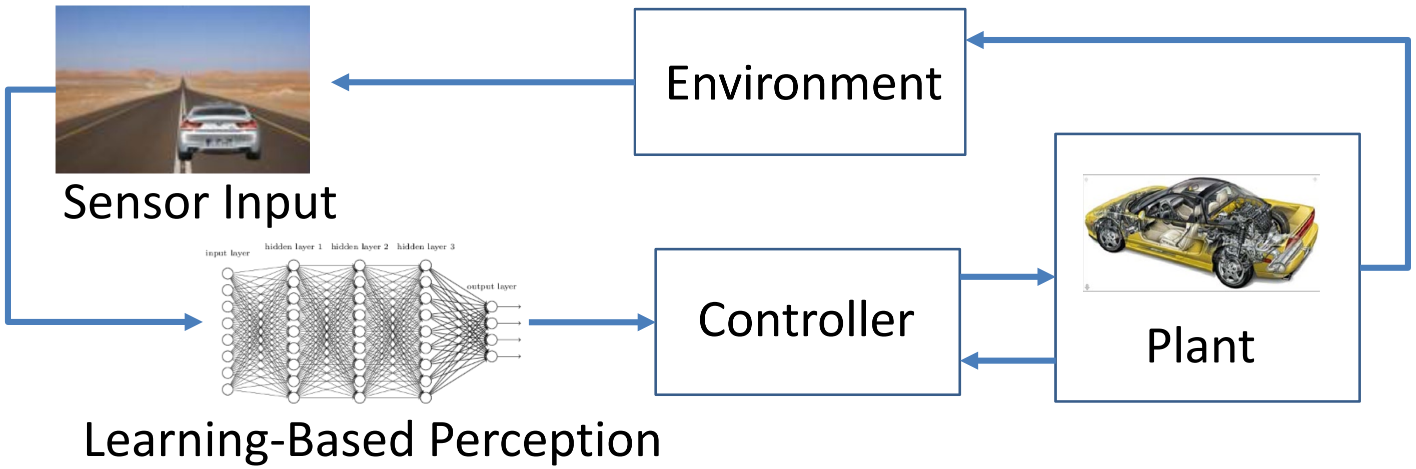
Large Scale Simulation

sity Dr
orial Way
Glade Path

Hardware Experiment

[A Scalable Framework for Real-time, Multi-Robot, Multi-Human Collision Avoidance, ICRA 2019]

A Semantic Approach to the Design of High-Assurance Learning-Based CPS



Sensor Input

Learning-Based Perception

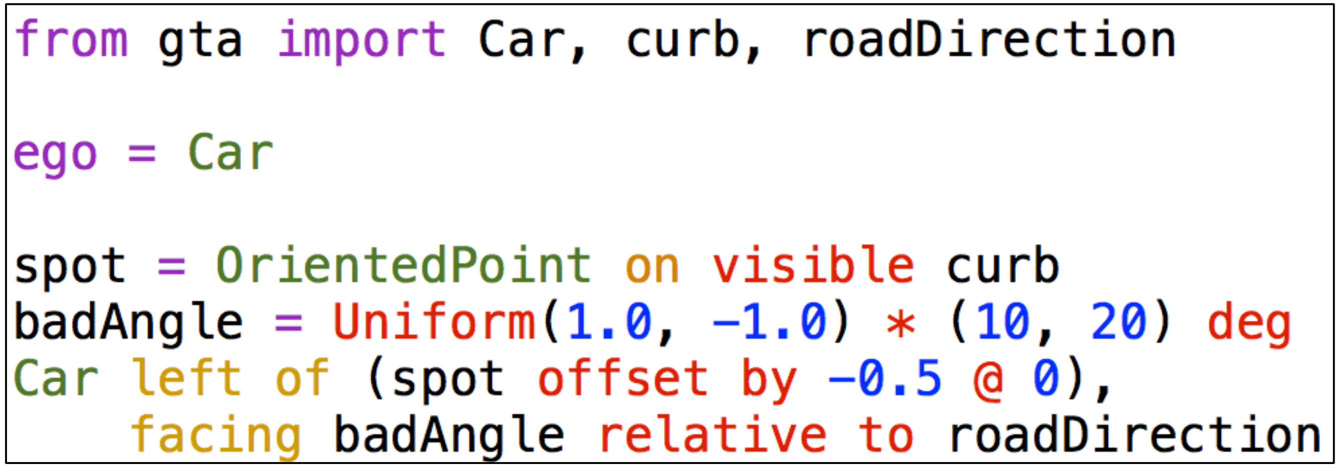
Environment

Controller

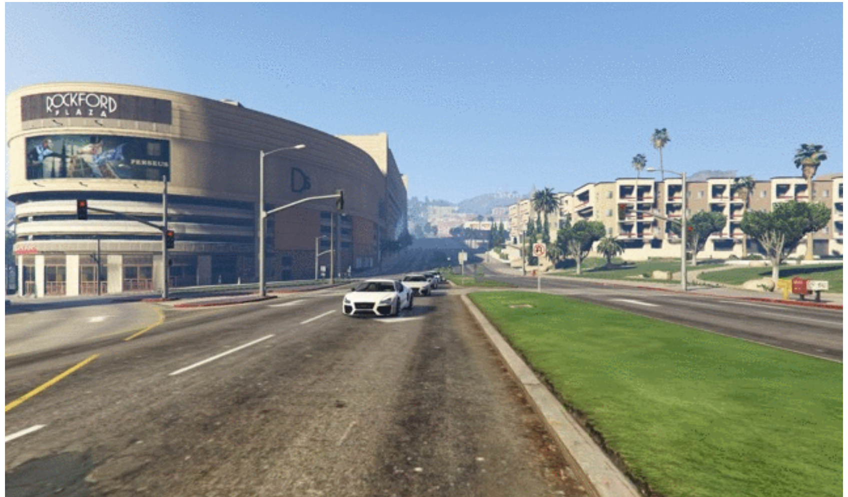
Plant

SCENIC: Scenario Description Language

- *Scenic* is a probabilistic programming language defining *distributions over scenes*
- *Use cases:* data generation, test generation, verification, debugging, design exploration, etc.

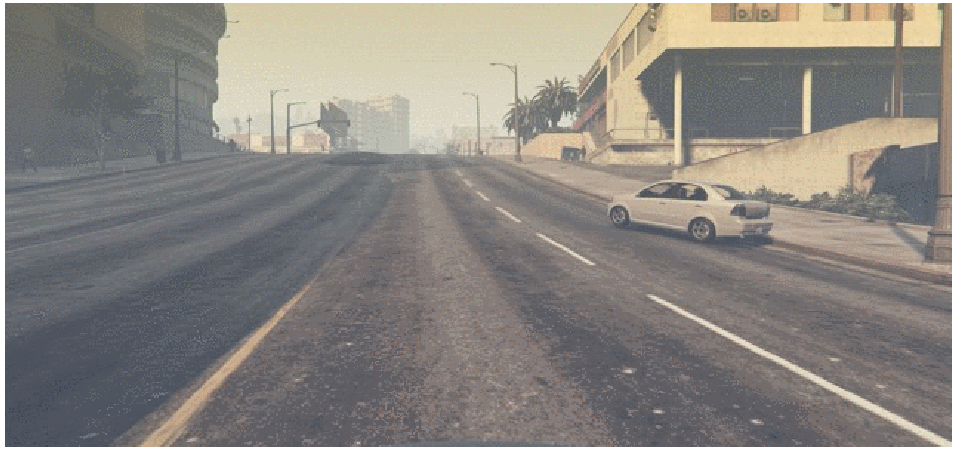


```
from gta import Car, curb, roadDirection

ego = Car

spot = OrientedPoint on visible curb
badAngle = Uniform(1.0, -1.0) * (10, 20) deg
Car left of (spot offset by -0.5 @ 0),
    facing badAngle relative to roadDirection
```

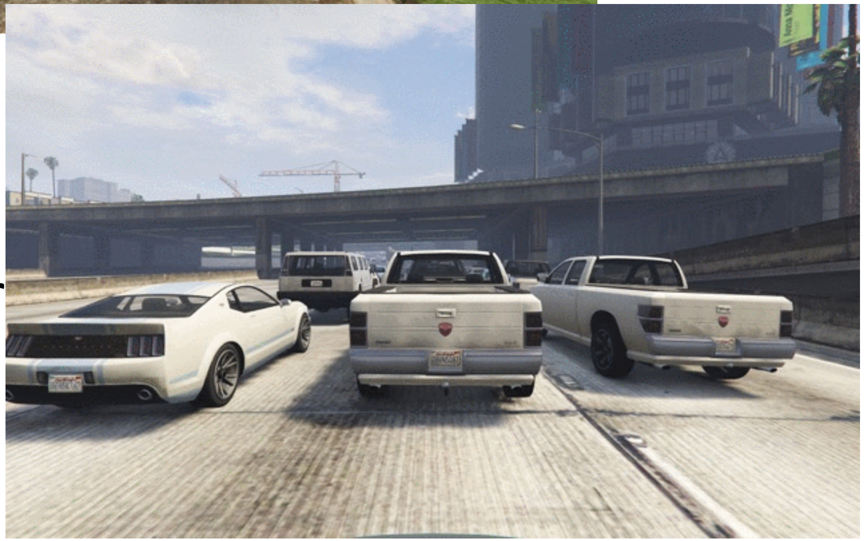


Platoons

Images created with GTA-V



Bumper-to-bumper



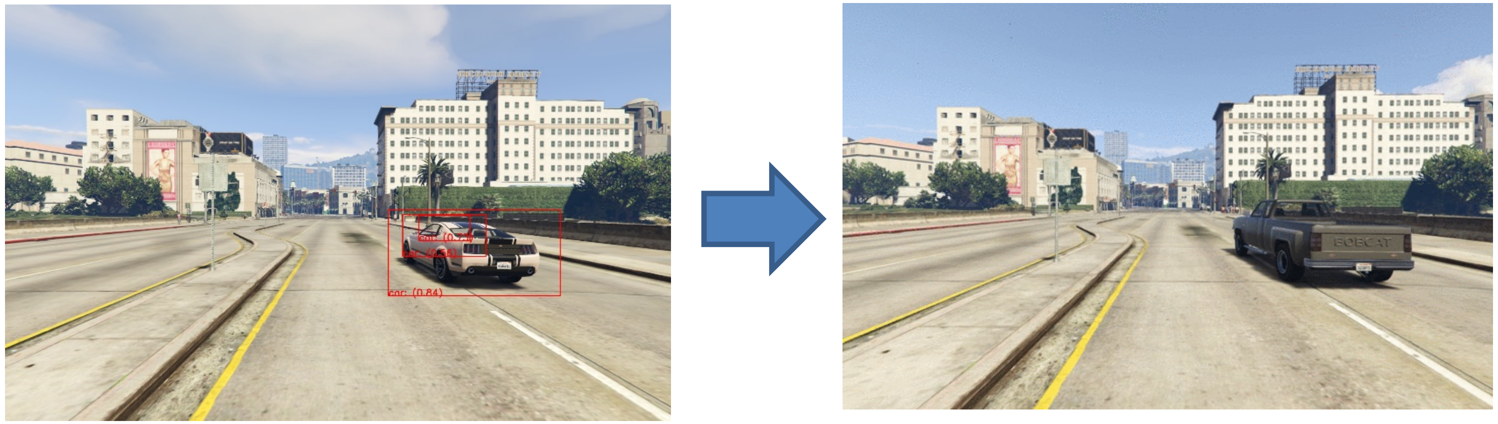
[D. Fremont et al., “Scenic: A Language for Scenario Specification and Scene Generation”, TR 2018, PLDI 2019.]

Some Applications of Scenic

- Data Generation, (Re)-Training
  - More controllable, interpretable
  - Improves performance significantly
  - Rare scenarios, controlled distributions, etc.

Car detection
with
occlusions

- Debugging Failures
  - Vary scenarios systematically
  - Explain failures of ML

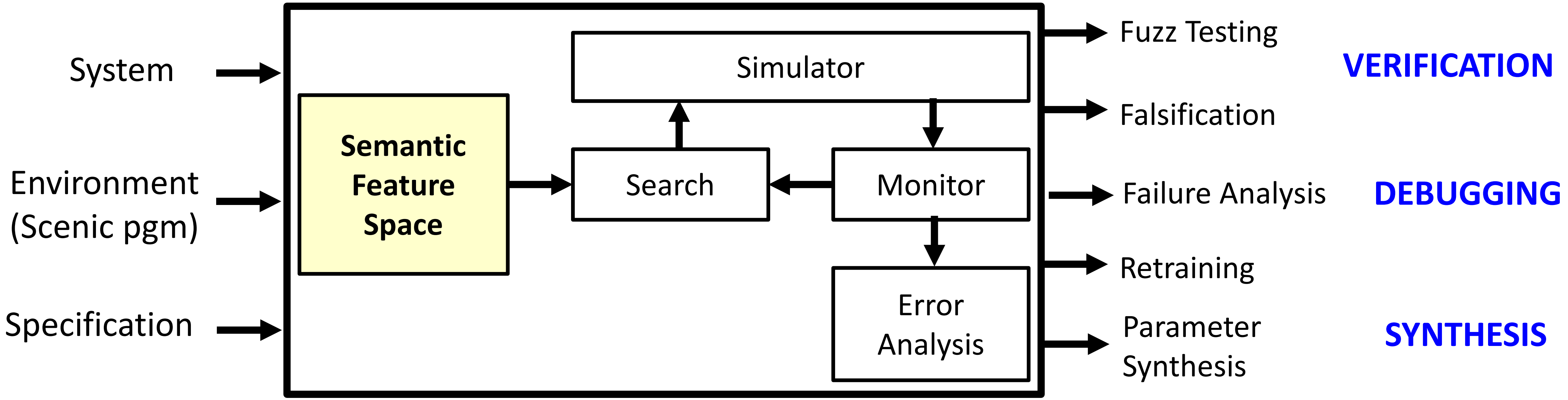


Test Hypothesis: does the car model lead to a mis-detection?

- Design Space Exploration

VERIFAI: A Toolkit for the Design and Analysis of AI-Based Systems [CAV 2019]

https://github.com/BerkeleyLearnVerify/VerifAI



System

Environment (Scenic pgm)

Specification

Semantic Feature Space

Simulator

Search

Monitor

Error Analysis

Fuzz Testing

Falsification

VERIFICATION

Failure Analysis

DEBUGGING

Retraining

Parameter Synthesis

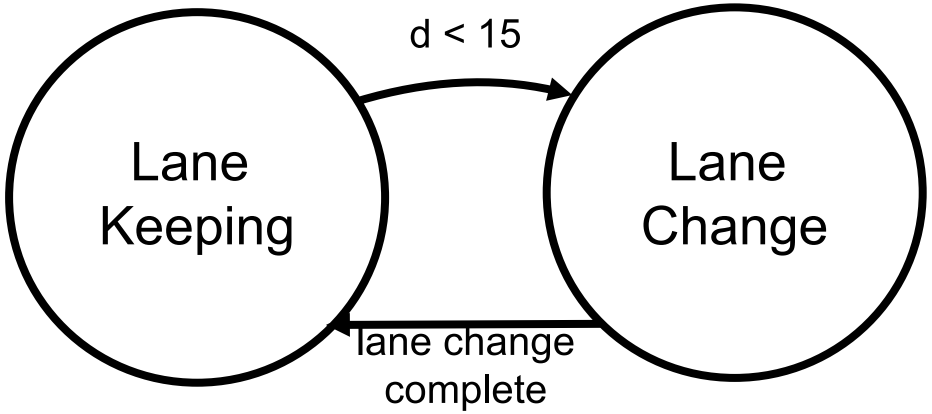
SYNTHESIS

ROBOTICS

AUTONOMOUS DRIVING

AIRCRAFT

Case Study for Temporal Logic Falsification with VerifAI: Navigation around an Accident Scenario

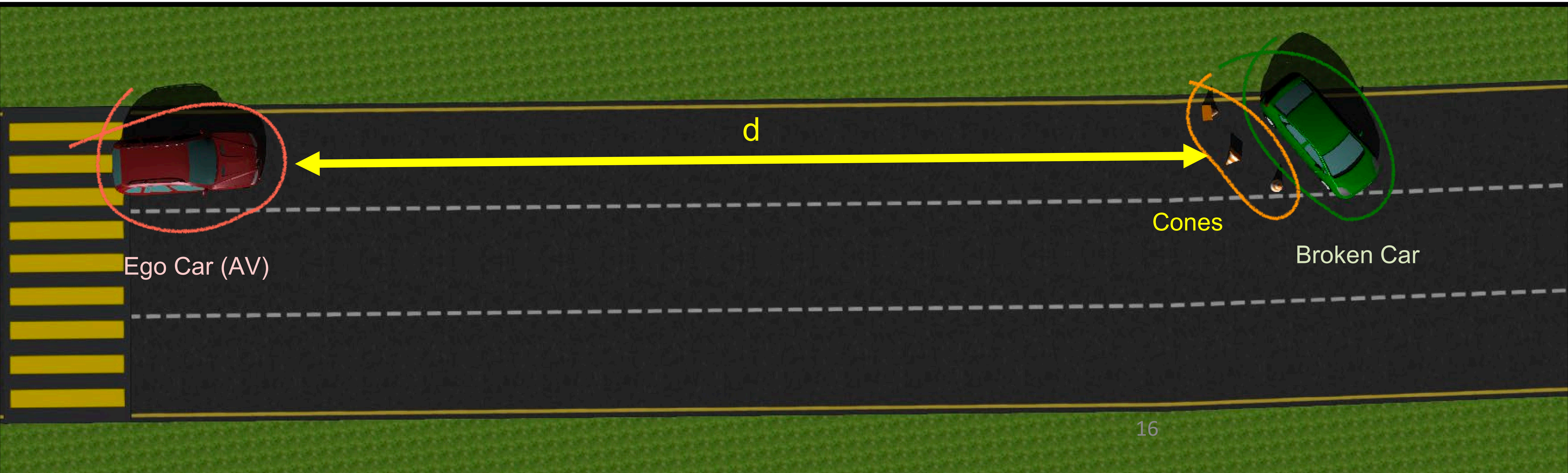


Lane Keeping

Lane Change

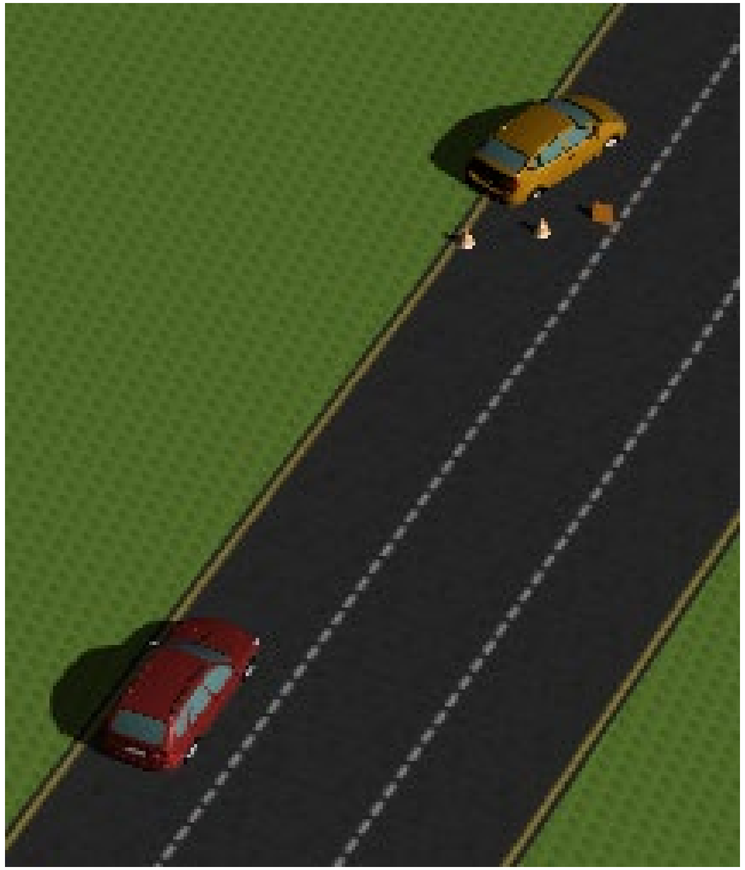
d < 15

lane change complete



Modeling Accident Scenario in the Scenic Language

```
# Pick location for blockage randomly along curb
blockageSite = OrientedPoint on curb

# Place traffic cones
spot1 = OrientedPoint left of blockageSite by (0.3, 1)
cone1 = TrafficCone at spot1,
                    facing (0, 360) deg


...

# Place disabled car ahead of cones
SmallCar ahead of spot2 by (-1, 0.5) @ (4, 10),
         facing (0, 360) deg
```



Fremont et al., *Scenic: A Language for Scenario Specification and Scene Generation*, PLDI 2019.

Temporal Logic Falsification

From Models to Real World: Bridging the Gap

SPE

Models

Real World

[HSCC 2019]

The

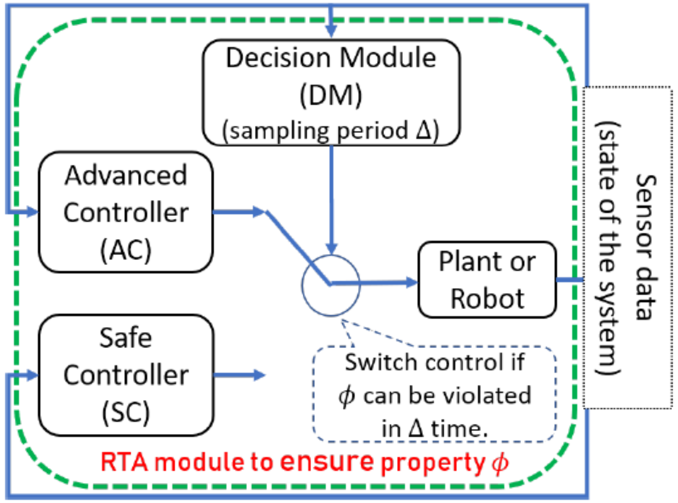
then

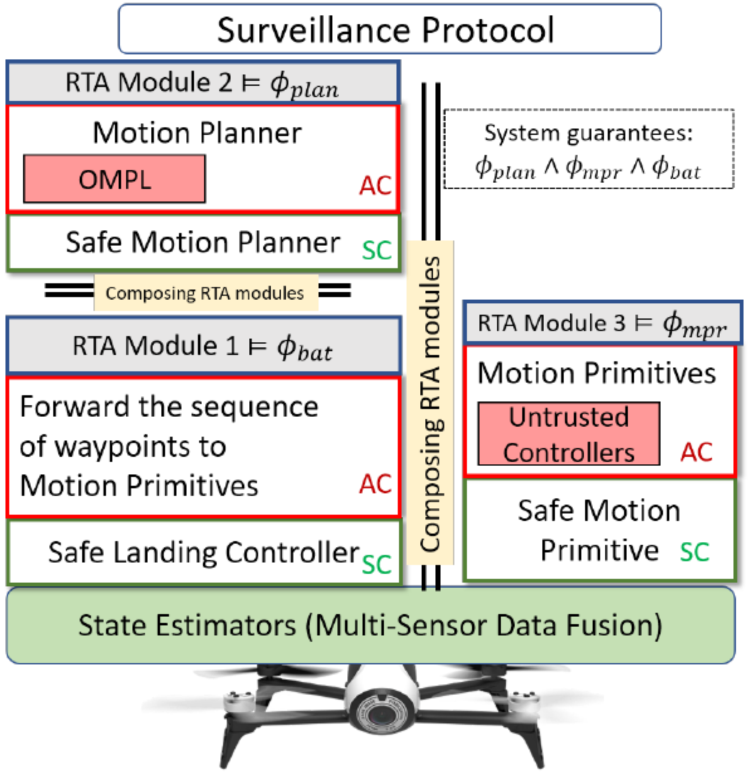
There

Mo

SOTER: Programming Framework for Run-Time Assurance

[DSN 2019]





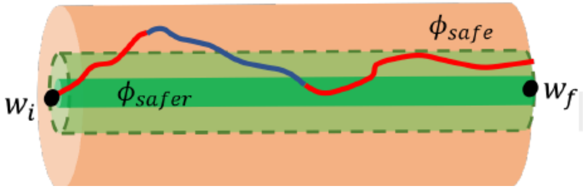
Surveillance Protocol
RTA Module 2 ⊨ ϕ_{plan}
Motion Planner
OMPL
AC
Safe Motion Planner
SC
Composing RTA modules
System guarantees:
$\phi_{plan} \wedge \phi_{mpr} \wedge \phi_{bat}$
RTA Module 1 ⊨ ϕ_{bat}
Forward the sequence of waypoints to Motion Primitives
AC
Safe Landing Controller
SC
Composing RTA modules
RTA Module 3 ⊨ ϕ_{mpr}
Motion Primitives
Untrusted Controllers
AC
Safe Motion Primitive
SC
State Estimators (Multi-Sensor Data Fusion)

Theorem: The following is an inductive invariant:

$$\text{Mode} = \text{SC} \wedge s \in \varphi_{safe}$$

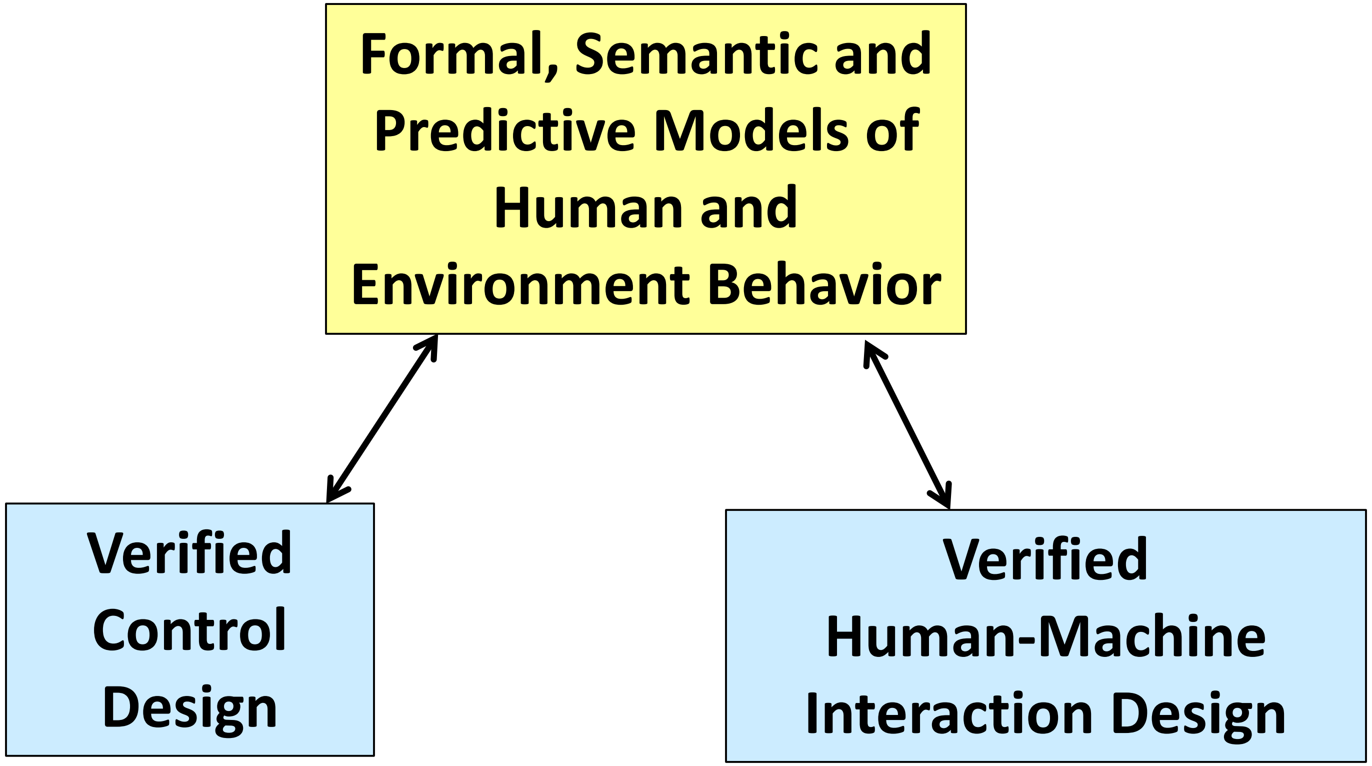
$$\vee$$

$$\text{Mode} = \text{AC} \wedge \text{Reach}(s,*,\Delta) \in \varphi_{safe}$$



```
module system = RTAModule1 || RTAModule2 || RTAModule3;
```

Formal Models Key to Co-Design



Formal, Semantic and Predictive Models of Human and Environment Behavior

Verified Control Design

Verified Human-Machine Interaction Design

A Selection of Other Results from VeHICaL

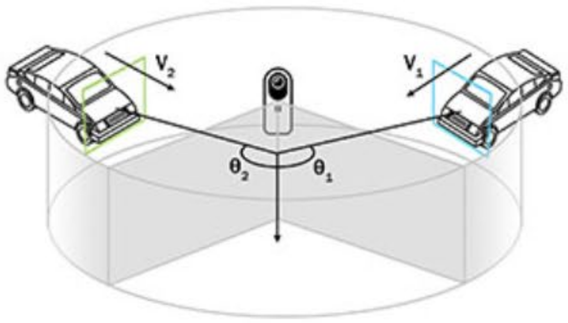
Come see the posters!

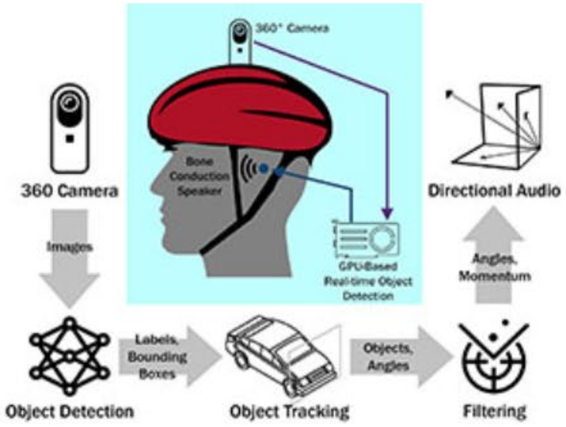
Abstractions for Neural Network Analysis

Vibro-Acoustical Approach to Driver Interfaces

Drowsy Driver Detection

HindSight: Bicyclist Assistance Systems





HindSight increases the environmental awareness of cyclists by warning them of vehicles approaching from outside their visual field. A panoramic camera mounted on a bicycle helmet streams real-time, 360-degree video to a laptop running YOLOv2, a neural object detector designed for real-time use. Detected vehicles are passed through a filter bank to find the most relevant.

Industrial Impact

- Several workshops with strong industry participation
- Open-Source Tools and Datasets
  - VerifAI, Scenic, …
  - Drowsy Driver Dataset, Visual-Acoustic Vehicle Dataset, …
- Tools/ideas being adopted by Industry

Ford Honeywell DENSO

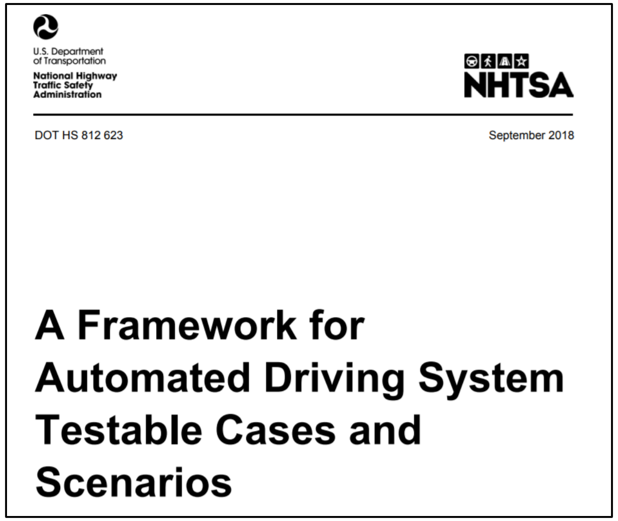
WAYMO United Technologies AAA

TOYOTA cruise LG

NISSAN nuTonomy ZOOX nuro

- Working with AAA & LG on AV scenario specification and testing at GoMentum test facility

- Advice to NHTSA project on AV Test Cases



U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA

DOT HS 812 623

September 2018

A Framework for Automated Driving System Testable Cases and Scenarios

Impact on Graduate and Undergraduate Education

- Several courses impacted by VeHICaL
- Reimagining Mobility – collaboration with Ford Greenfield Labs
  - at the Jacobs Institute of Design Innovation @ Berkeley
- Academic/industry positions for graduates from VeHICaL project

UIUC

Stanford

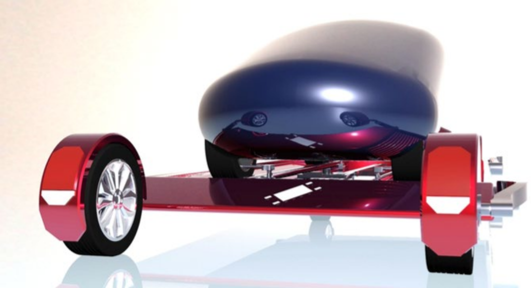
UCSC

Princeton

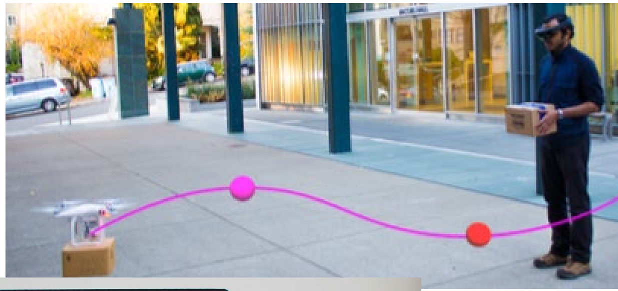
TRI

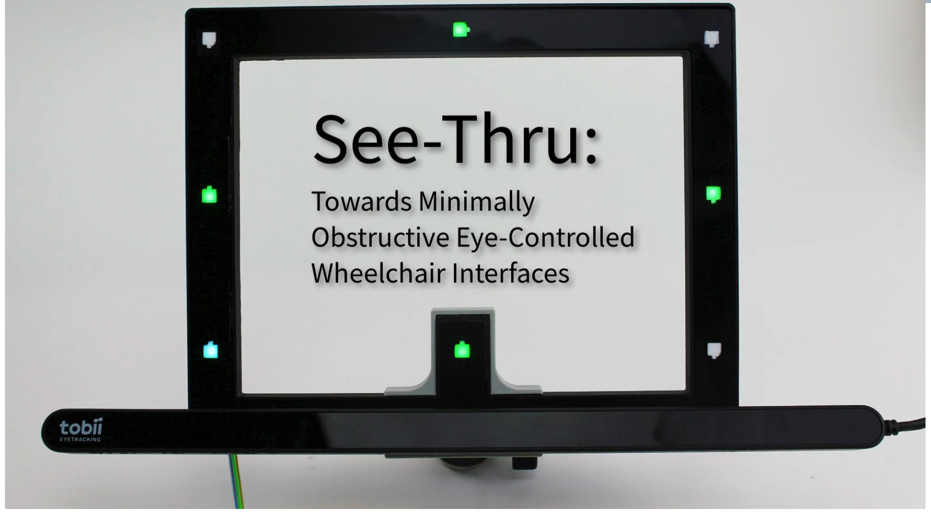
Waymo

… and more



Design Project 3:
The Goods Delivery Interface
Between Humans and
Autonomous Vehicles

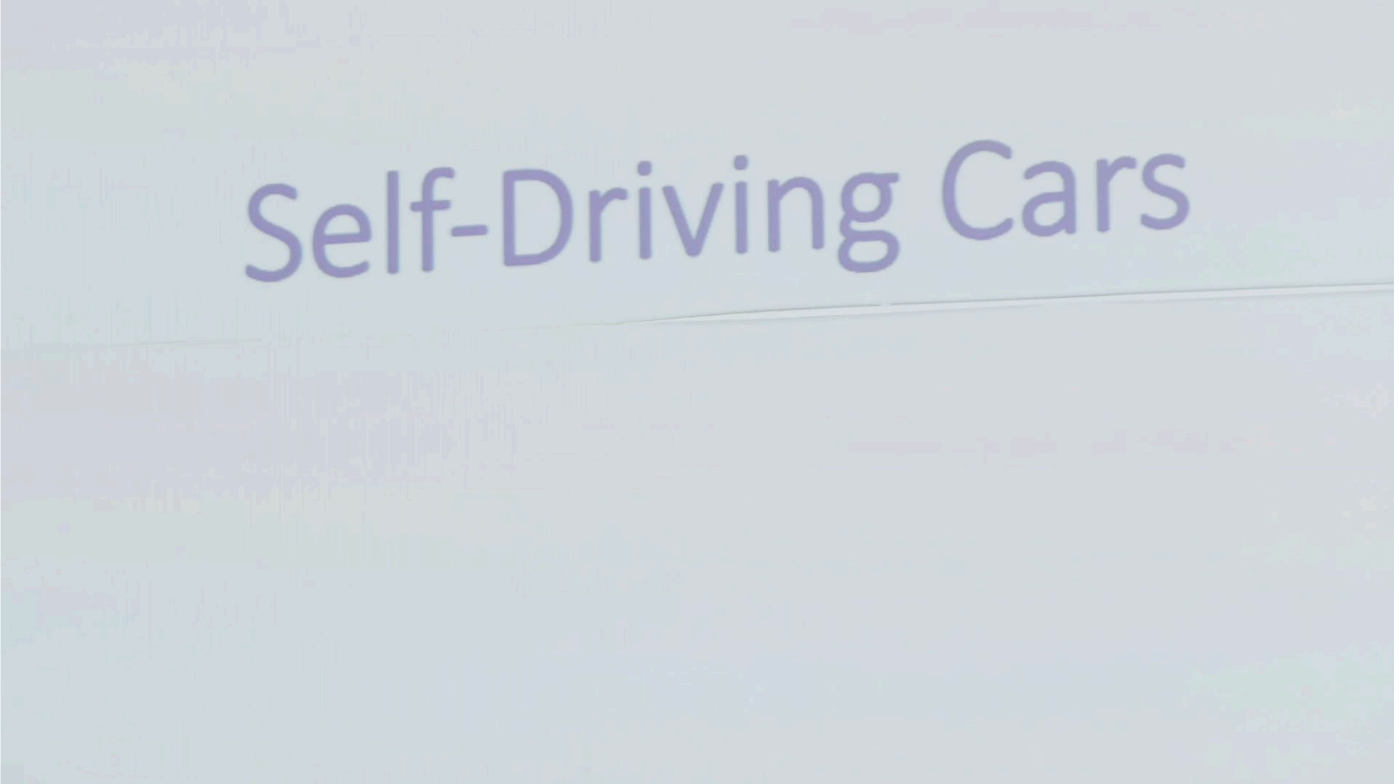




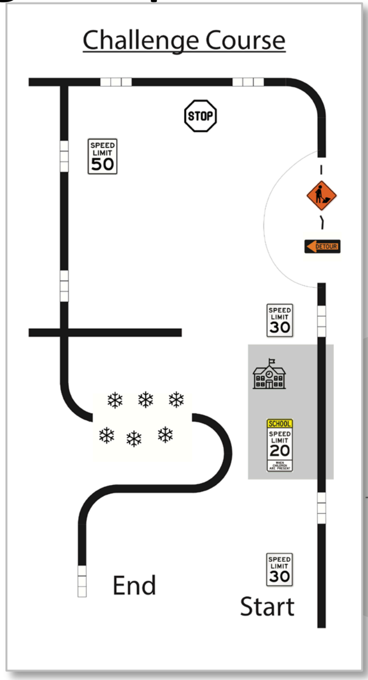
See-Thru:
Towards Minimally Obstructive Eye-Controlled Wheelchair Interfaces

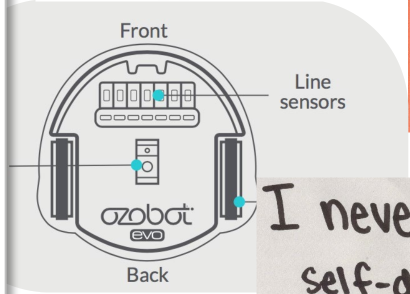
Broader Impacts – Girls in Engineering (GiE) VeHICaL modules

- Summer program for middle-school girls at Berkeley
- VeHICaL provided instructors/mentors, funding, content
- Modules on self-driving car technology using simple Ozobot platform



Self-Driving Cars





-I-
never knew:
ANYTHING ABOUT
SELF-DRIVING
CARS

I never knew that...
self-driving cars
were so hard
to make.

I never knew...

Before this camp, I never knew about lidar and radar. I thought self-driving cars just used cameras to sense.

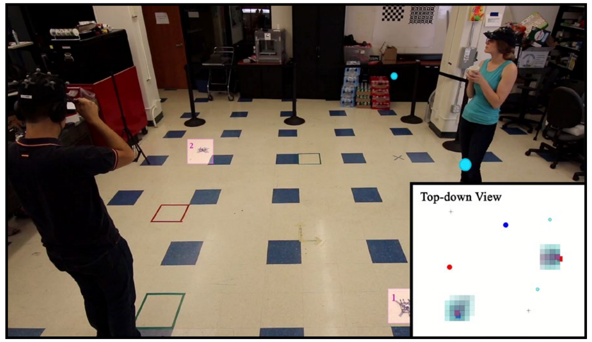
VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems

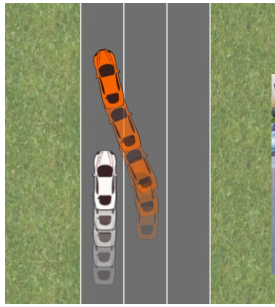
Challenge:

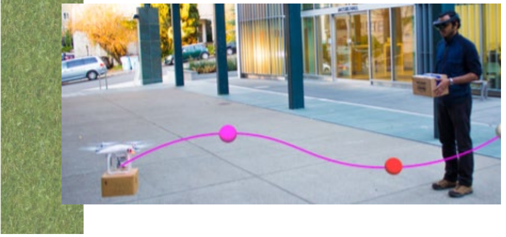
- *Co-design human interfaces and control* for human-cyber-physical systems with *provable guarantees*
- Apply to semi-autonomous vehicles (ground and air)

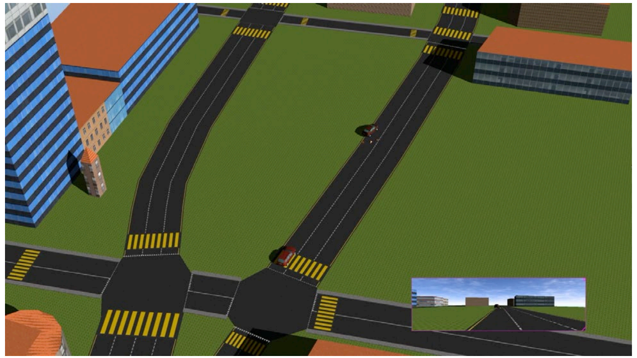
Solution:

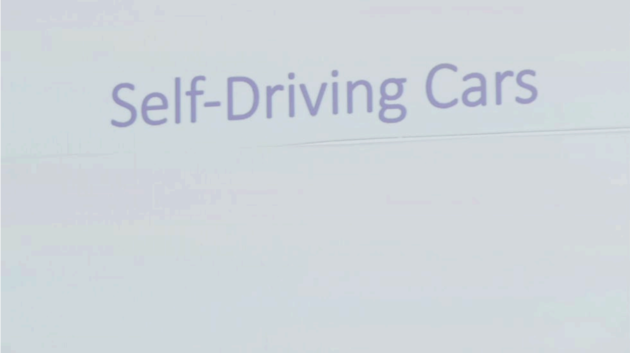
- Integrate Learning, Verification and Control
- Data-Driven Resource Rational Human Modeling
- Prototype Controllers & Interfaces, Evaluate on Testbed

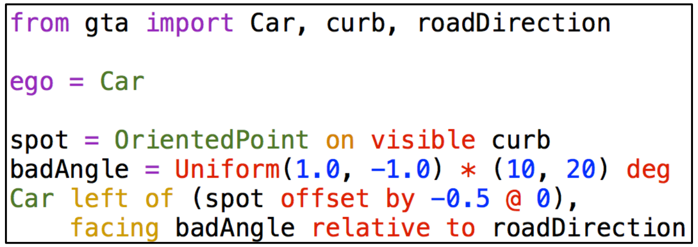












```
from gta import Car, curb, roadDirection

ego = Car

spot = OrientedPoint on visible curb
badAngle = Uniform(1.0, -1.0) * (10, 20) deg
Car left of (spot offset by -0.5 @ 0),
    facing badAngle relative to roadDirection
```

Scientific Impact:

- Developing a Science of Co-Design of Human Interfaces and Control
- Bridging Model-Based and Data-Driven Design of CPS

Broader Impact:

- Significantly improve safety, security, and performance of systems where humans interact closely with automation
- Involve middle/high-school and undergraduate students in VeHICaL activities

THANK YOU!

CPS Awards 1545126, 1544714, 1544924

S. A. Seshia, R. Bajcsy, B. Hartmann, S. S. Sastry, C. Tomlin (UC Berkeley), T. Griffiths (Princeton), R. Murray (Caltech), C. Sturton (UNC).