

Adjudication of Symbolic & Connectionist Arguments in Autonomous Driving AI

Michael Giancola
Selmer Bringsjord
Naveen Sundar Govindarajulu
John Licato



6th Global Conference on Artificial Intelligence (GCAI 2020)
October 30th, 2020



Sponsored by

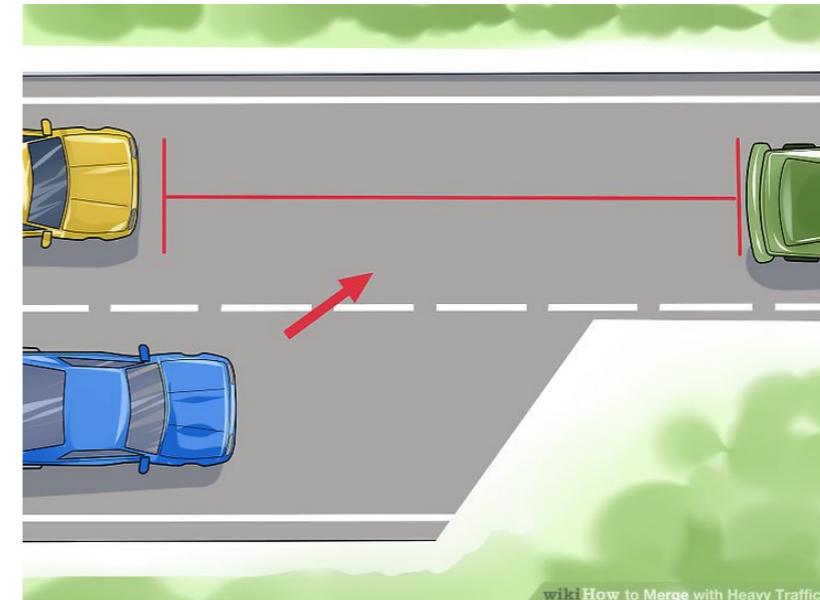


Advocating for a Hybrid Approach to Autonomous Driving

Machine Learning



Reasoning w/ Intensional Content

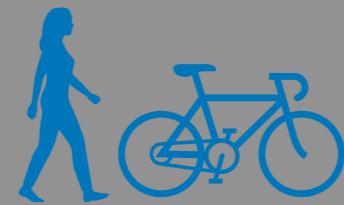


A Tragic Case Study

- March 18, 2018:
The first pedestrian was killed by an autonomous car.
- Uber's ATG operator was distracted (watching TV on phone).
- However, the accident was directly caused by a flaw in Uber's driving system.
- We argue that by integrating symbolic methods — in particular, formal argumentation — with the connectionist methods currently employed, the accident could've been avoided.



At the start of the event, the vehicle was in autonomous mode in the rightmost of four lanes traveling in the same direction, and the pedestrian was walking her bicycle across the street starting on the leftmost side of the roadway.



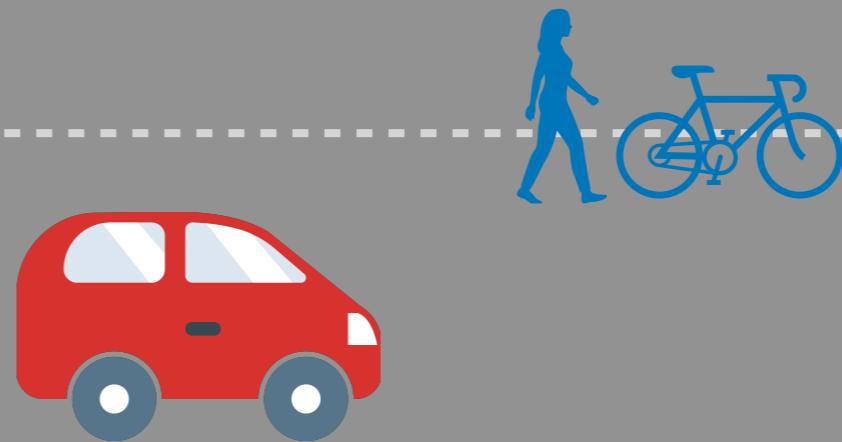
The vehicle's radar first detected the pedestrian **5.6 seconds** before the fatal collision. Less than half a second later, the lidar detected the pedestrian but classified her as "**Other**".



For the next 2.5 seconds, the lidar re-classified her several times, alternating between “Vehicle” and “Other”. The vehicle’s automated-driving system (ADS) attempted to predict her direction of travel several times, but **discarded any previous information about her trajectory every time it reclassified her**.



With 2.6 seconds until collision, the lidar classified her as a bicycle but, as it was yet again changing her classification, discarded any past trajectory information, and hence determined that she was not moving. Up to this point, **the car had not taken any evasive or corrective action.**



With 1.5 seconds left, the lidar re-classified her yet again, this time as “Unknown”. **The system once again loses all of its tracking history.** However, since at this point the pedestrian had entered the vehicle’s lane, the ADS generated a plan to turn the car to the right to avoid her.

Now, how our AI
would've handled it...



Three hundred milliseconds later, the lidar re-classified her as a bicycle, and determined that it would be impossible at this point to maneuver around her. With just 200 ms until collision, the ADS began braking the vehicle, pitifully too late to stop in time.

Argument 1

$P(r_1, t_0, At(o^*, t_0, \ell_1))$

Argument 2

$P(r_2, t_0, At(o^*, t_0, \ell_1))$



Label the time point at which the radar first detected the pedestrian t_0 ,
and the location of the pedestrian at that time ℓ_1 .

Argument 1

$P(r_1, t_0, At(o^*, t_0, \ell_1))$

$P(r_1, t_1, At(o^*, t_1, \ell_1))$

...

$B(r_1, t_1, \neg Moving(o^*))$

\succ_t^a

$B(r_1, t_1, Moving(o^*))$

$\therefore B^2(r_1, t_1, \neg Moving(o^*))$

$\therefore B^2(r_1, t_1, \neg NeedToBrake(c))$

Argument 2

$P(r_2, t_0, At(o^*, t_0, \ell_1))$

$P(r_2, t_1, At(o^*, t_1, \ell_2))$

$P(r_2, t_1, \ell_1 \neq \ell_2)$

...



$\therefore B^5(r_2, t_1, Moving(o^*))$

$\therefore B^5(r_2, t_1, NeedToBrake(c))$

Adjudicator

$B(a, t_1, NeedToBrake(c))$



Then, denote the next time it detects and re-classifies
the pedestrian as t_1 , and new location ℓ_2 .

Argument 1

$P(r_1, t_0, At(o^*, t_0, \ell_1))$

$P(r_1, t_1, At(o^*, t_1, \ell_1))$

...

$B(r_1, t_1, \neg Moving(o^*))$

\succ_t^a

$B(r_1, t_1, Moving(o^*))$

$\therefore B^2(r_1, t_1, \neg Moving(o^*))$

$\therefore B^2(r_1, t_1, \neg NeedToBrake(c))$

Argument 2

$P(r_2, t_0, At(o^*, t_0, \ell_1))$

$P(r_2, t_1, At(o^*, t_1, \ell_2))$

$P(r_2, t_1, \ell_1 \neq \ell_2)$

...

$\therefore B^5(r_2, t_1, Moving(o^*))$

$\therefore B^5(r_2, t_1, NeedToBrake(c))$

Adjudicator

$B(a, t_1, NeedToBrake(c))$



Next Steps

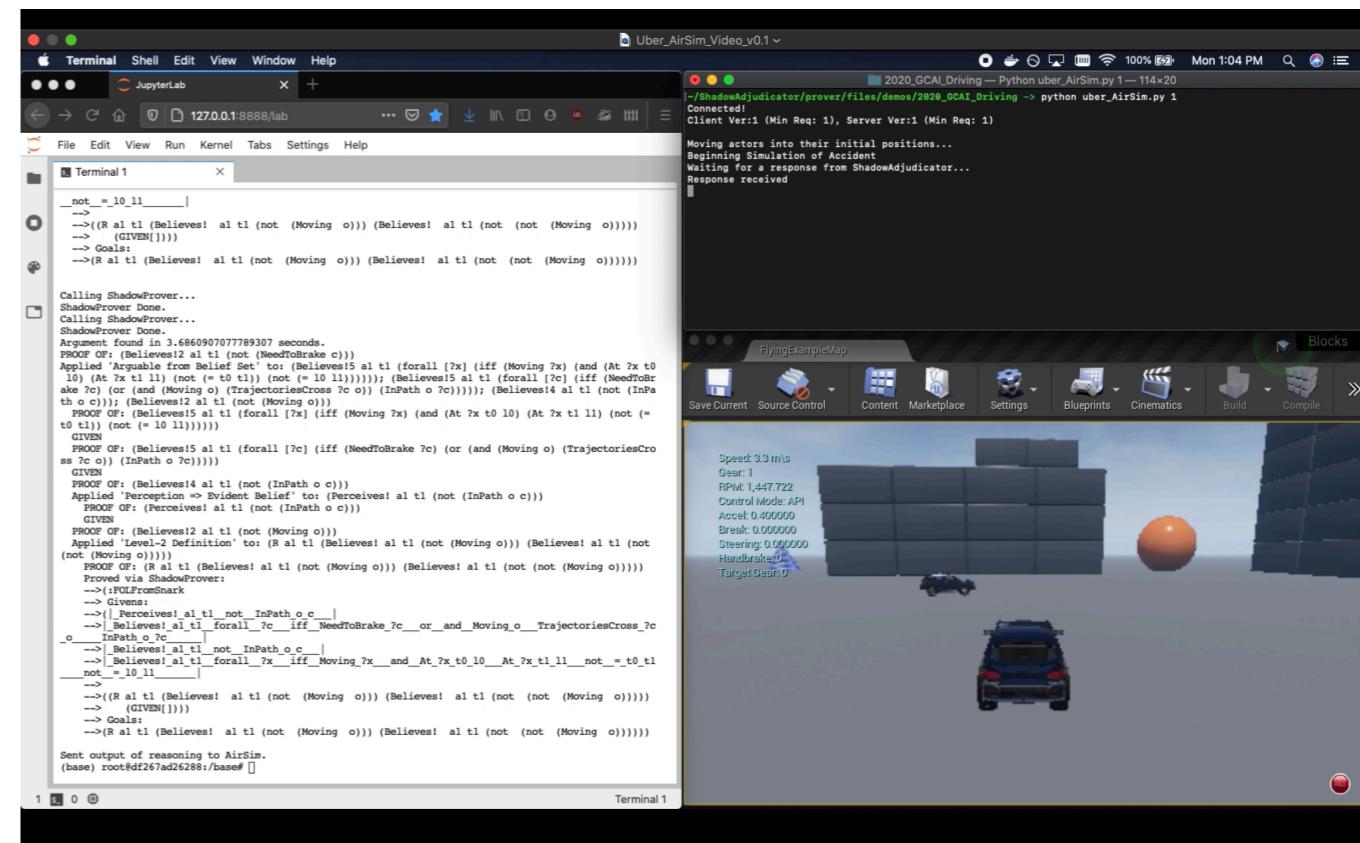
- Since Publication

- Fully formalize arguments
- Create an automated reasoner which can find the arguments

- Future

- Implement a simulation using Microsoft AirSim

- Publish full-length paper



Questions?