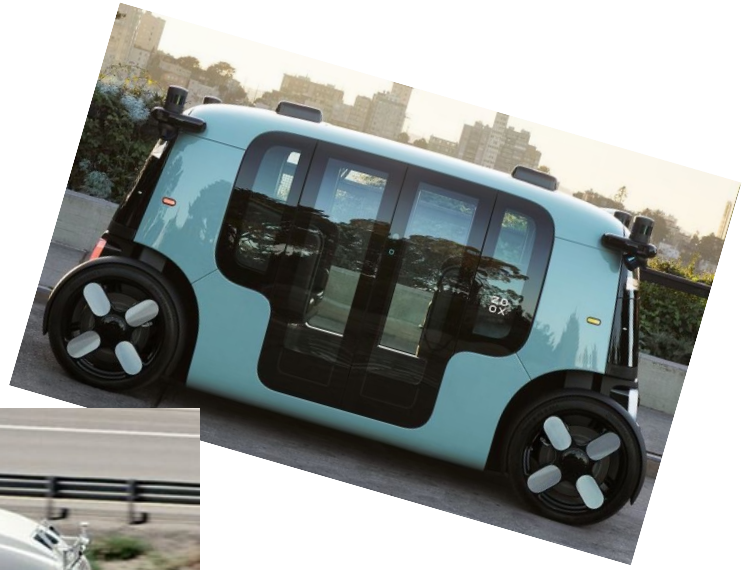
A Mechanism to Assess the Safety of (Semi-)Autonomous Cars

Hanna Kurniawati

In collaboration with Jimy Cai Huang

Australian National University | RESEARCH SCHOOL OF COMPUTER SCIENCE

Lots of Work on Autonomous Cars



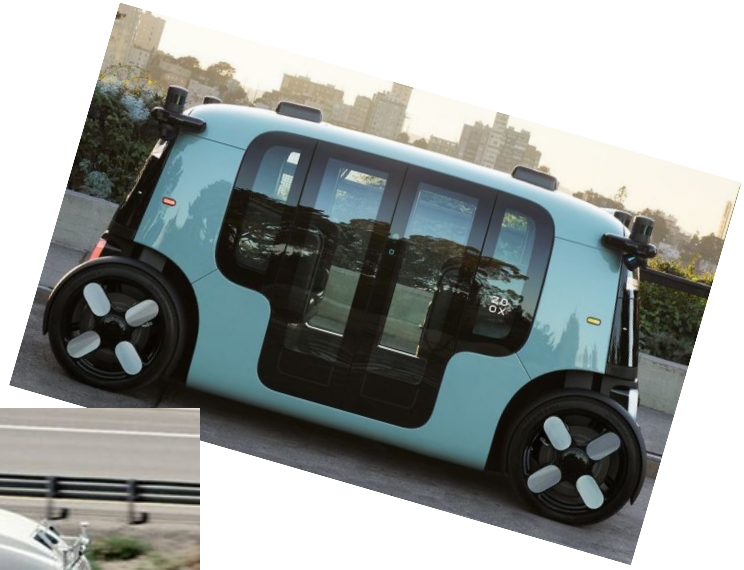
Pictures taken from news release of each car

Lots of Discussions on Safety

- Definitions of safety
- AI (incl. ML) with safety guarantees
- Formal-method based verification
- Finding “bugs” (aka. scenarios that cause accidents)

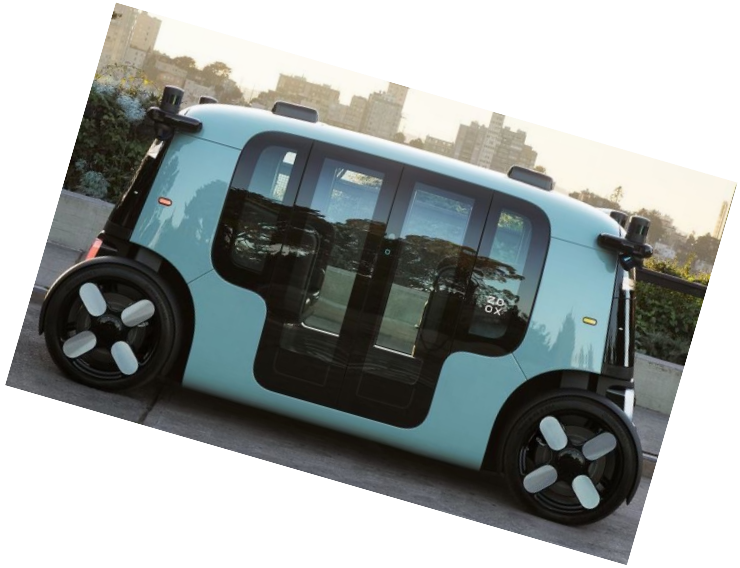
From and for developers

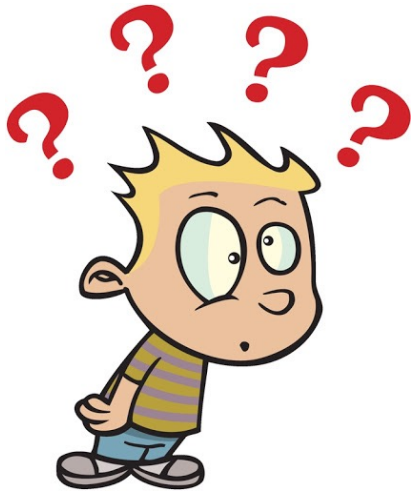
How about consumers?



Pictures taken from news release of each car

How about consumers?

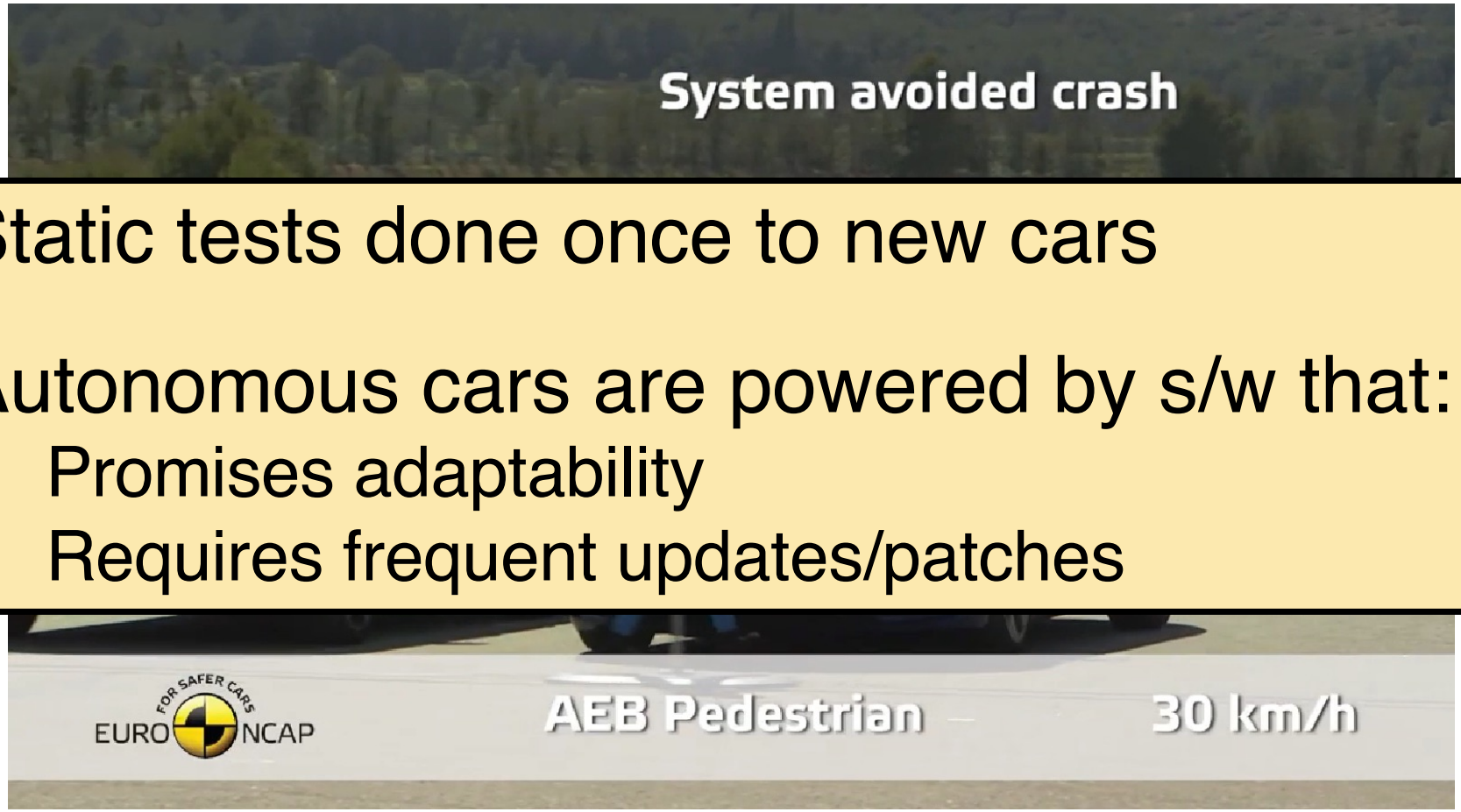




Pictures taken from news release of each car and http://clipart-library.com

In terms of Safety?

- NCAP (New Car Assessment Programme)



Static tests done once to new cars

Autonomous cars are powered by s/w that:

- Promises adaptability
- Requires frequent updates/patches

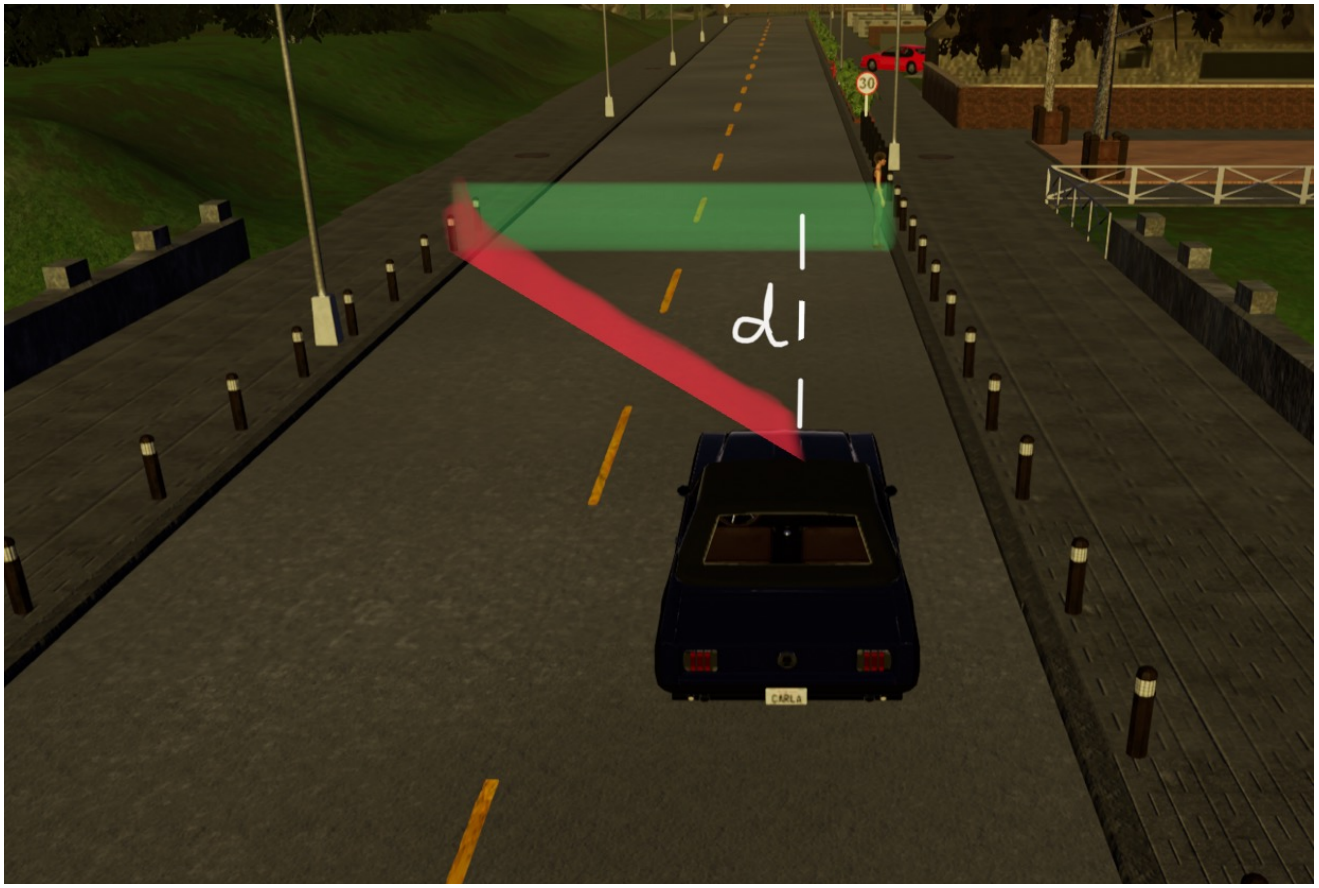
Our Proposal

Picture taken from the facebook page of starcarwash

Desirable Properties

- A simple safety indicator
  - One number, easy to understand by users
- Fast and easy assessment mechanism
  - Assessment can be done frequently (e.g., after every s/w updates, perhaps imposed during registration renewal)

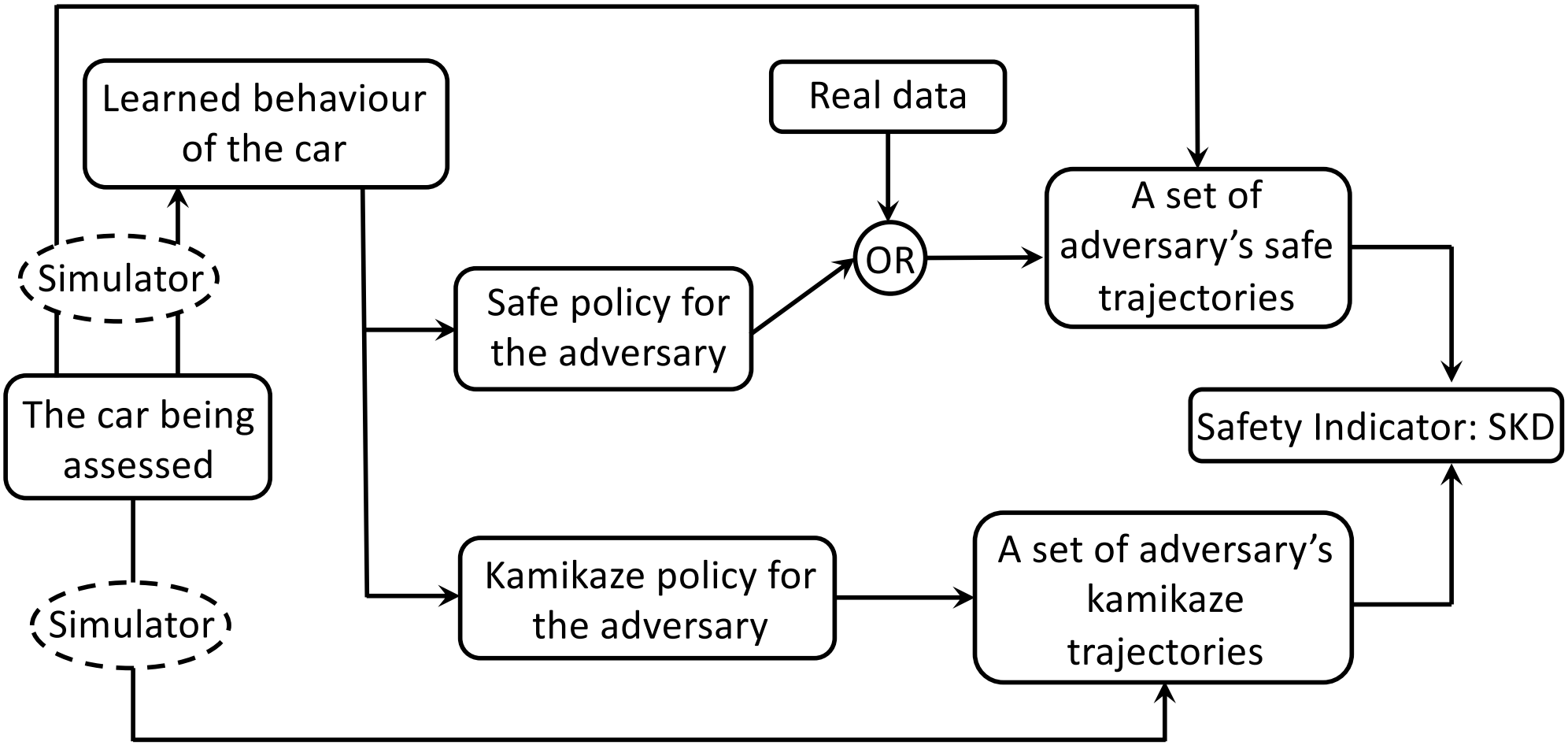
The Hypothesis



The distance between safe and dangerous trajectories of the adversary may serve as a safety indicator.

Distance computation is generally fast!

Proposed Mechanism



Dashed ellipse: May not be used

Safe Trajectories

- Real data
- Provided by regulatory body
- Generate synthetic data
  - Strive for optimal to get to the destination [biological hypothesis, Breed & Moore'15]
  - Account for the car's behaviour, though don't know their exact policy → partially observed
  - Use Partially Observable Markov Decision Process (POMDP)
- A trajectory:
  - A mapping from [0, 1] to position of the adversary in the operating environment (in this case, a bounded R^2)
  - Can be approximated by a polygonal chain (a sequence of line segments)

Kamikaze Trajectories

- Given a safe adversary's trajectory ϕ
- Find a sampled set of adversary's trajectories closest to ϕ that causes collision with the car being tested
  - The exact policy of the car is not known in advance and need to be learned → car's behaviour is partially observed
  - Closest: Current practice, within certain distance away from ϕ
  - Use Partially Observable Markov Decision Process (POMDP)

The Safety Indicator: SKD

- Given:
  - A set of safe trajectories Φ
  - A set of kamikaze trajectories $\Psi(\phi)$ for each safe trajectory $\phi \in \Phi$
- Suppose $\Psi = \bigcup_{\phi \in \Phi} \Psi(\phi)$, then

$$SKD(\Phi, \Psi) = \frac{1}{|\Psi|} \sum_{\phi \in \Phi} \sum_{\psi \in \Psi(\phi)} d(\phi, \psi)$$

SKD = Safe-Kamikaze Distance

$d(\phi, \psi) = inf_{\alpha,\beta}\, max_{t \in [0,1]} \left\| \phi\big(\alpha(t)\big) - \psi(\beta(t)) \right\|$

Also known as Fréchet distance

Turns out …

If $SKD(\Phi, \Psi) = \delta$ and we have sufficiently many samples, then

$$Prob(d(\phi', \psi') < \eta) \leq 2\left(\frac{\eta}{\delta^2} + \frac{\eta}{\delta}\right)$$

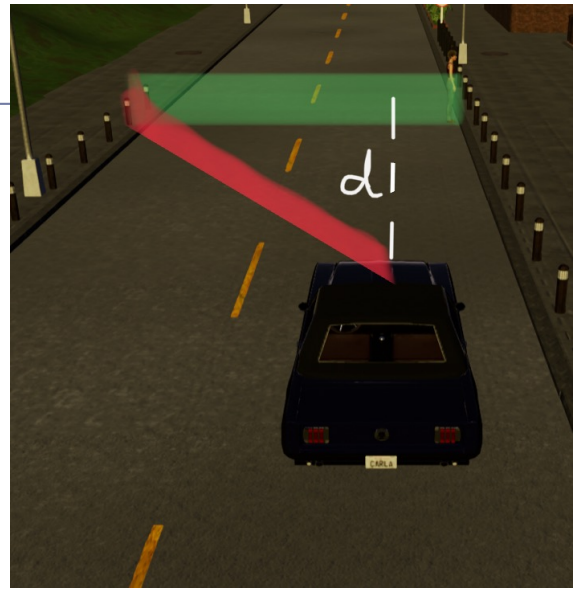
for small $\eta \in (0, \delta)$ and any safe trajectory ϕ' and corresponding dangerous trajectory ψ' sampled from the same distribution as the one used to generate Φ and Ψ

The probability that a small deformation changes a safe adversary's trajectory into a dangerous one is upper bounded by a value inversely proportional to SKD

So…

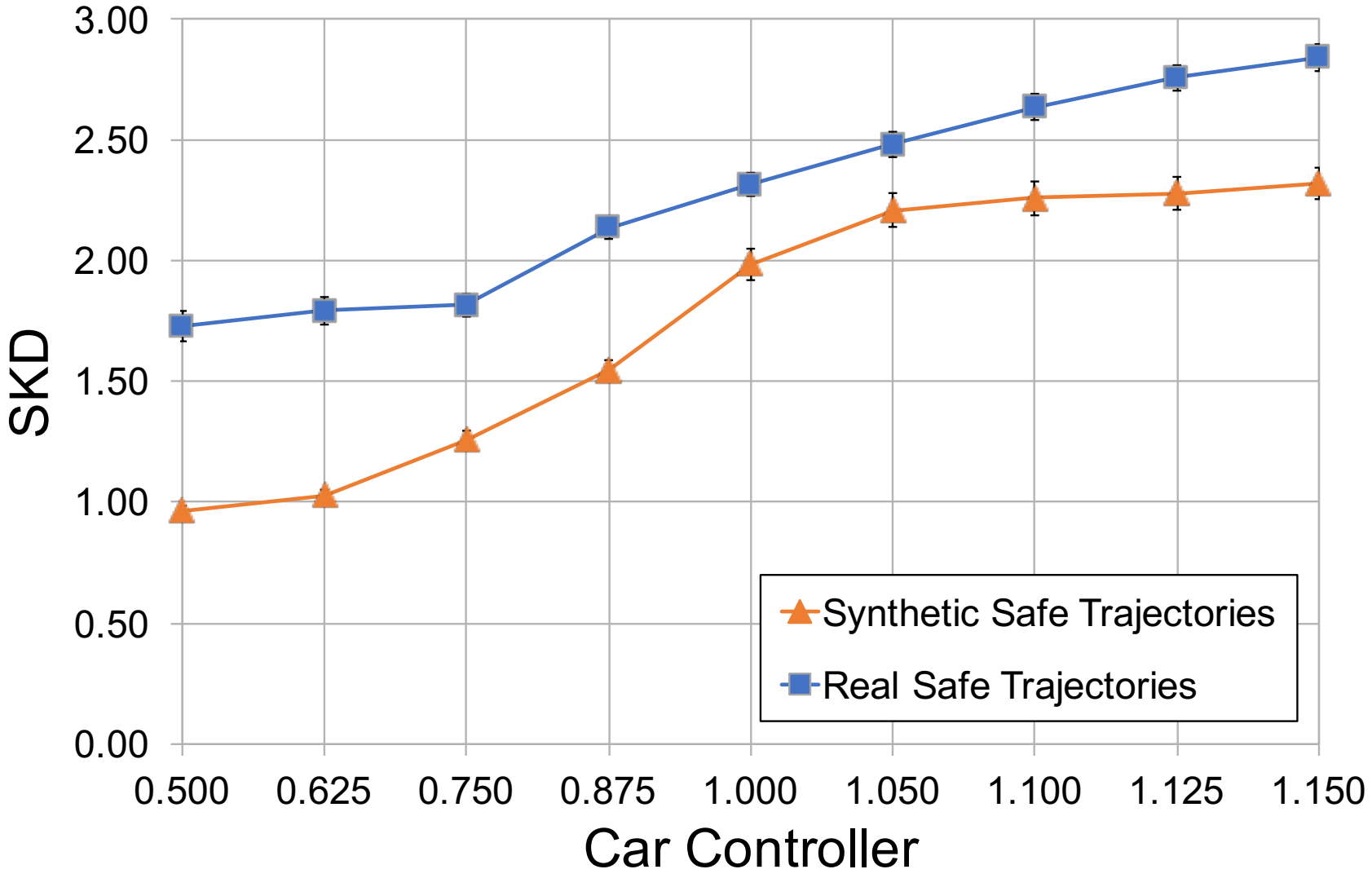
- Large SKD is a possible sufficient condition for an autonomous car to be safe. Meaning:
  - If SKD is large, the car is likely to be safe
  - If SKD is small, unfortunately, we can’t say much!!!

Systematic Test

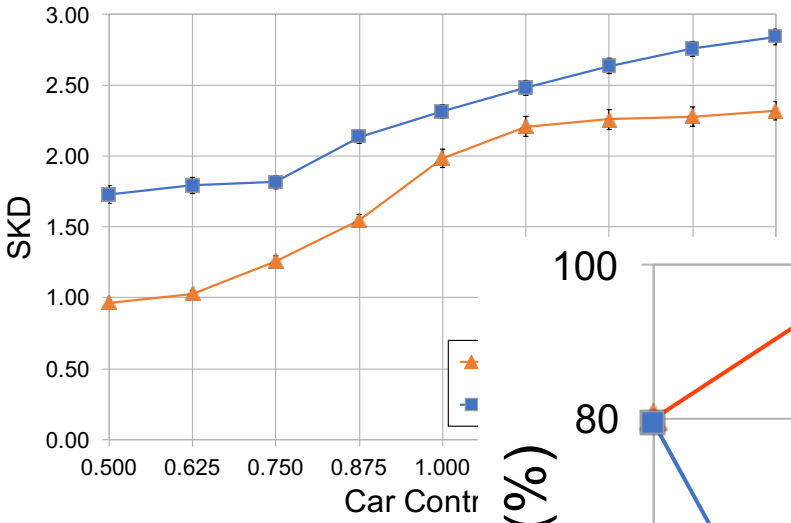


- Scenario: Euro-NCAP Vulnerable Road User test on Autonomous Emergency Breaking
- A pedestrian crossing a road in front of the car moving on a single lane
- The car:
  - Has a maximum velocity & fixed acceleration/deceleration
  - Safe stopping distance κ: The distance to move from maximum velocity to 0 with maximum deceleration
  - Where the car starts to decelerate: $C.\kappa$ distance away from the pedestrian
    - C: Multiplier, larger means safer

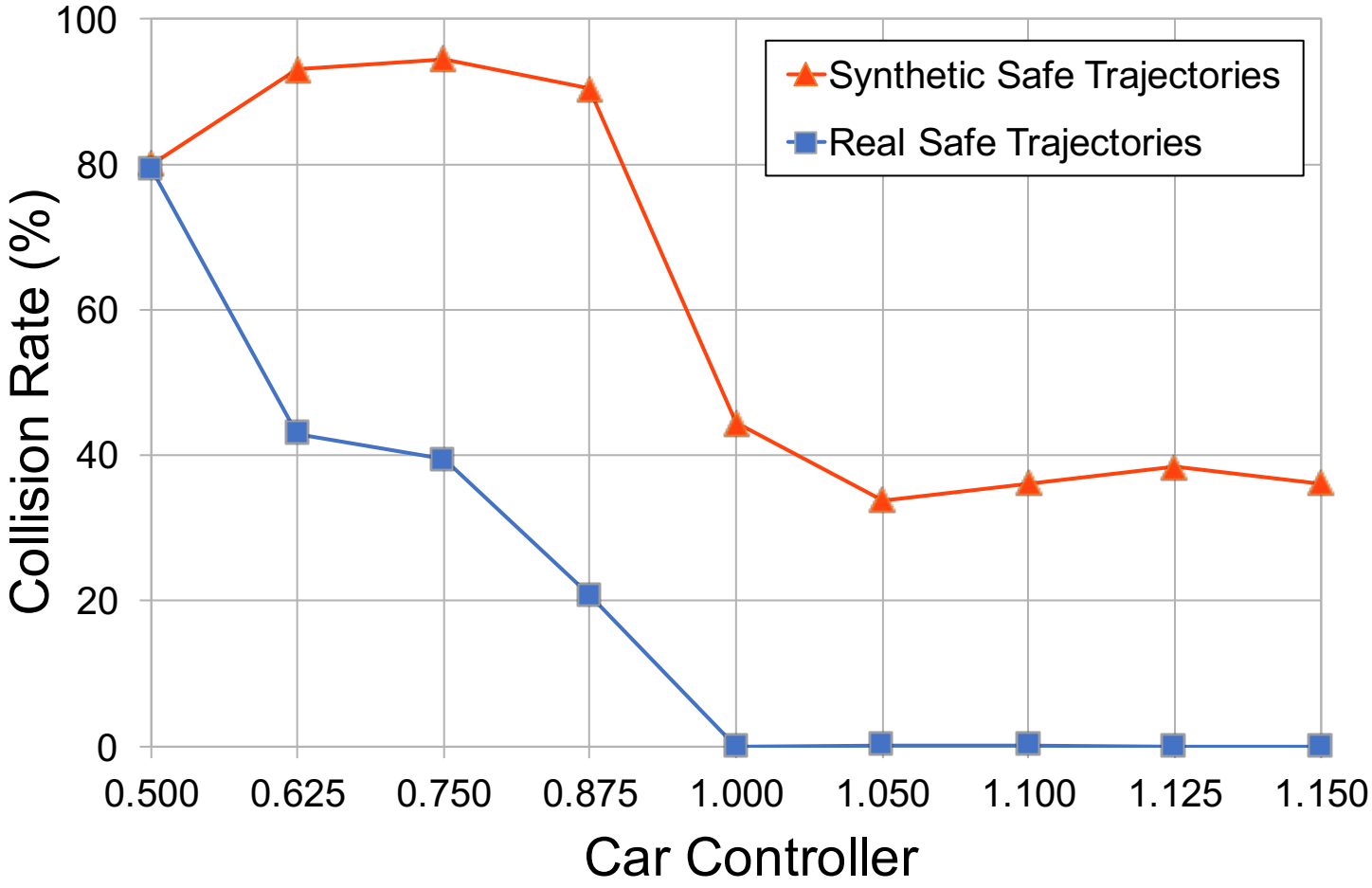
Simulation Results



Simulation Results

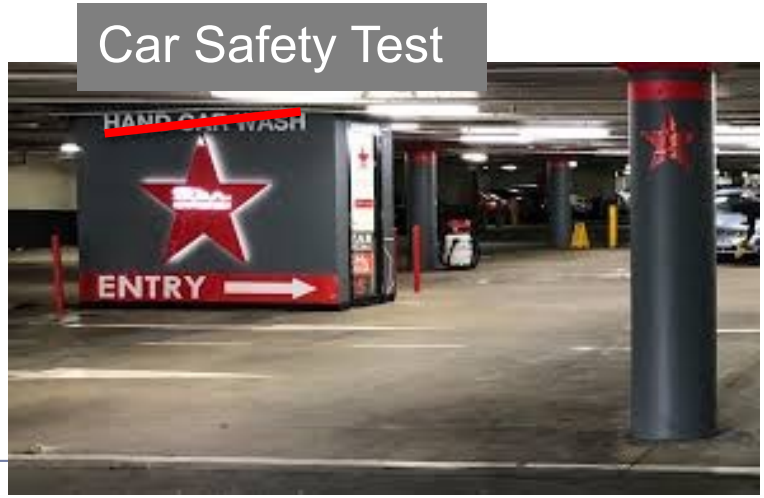


Average time to generate trajectories + computing SKD < 10s per run



Summary

- The safety indicator SKD can be used to upper bound the probability that a small deformation changes a safe adversary's trajectory into a dangerous one
- SKD with sufficient statistical confidence can be computed in under 30 minutes using a typical desktop (i7 quad-core)



Thank you

Q&A