AN OPEN, TRANSPARENT, INDUSTRY-DRIVEN APPROACH TO SAFETY

Jack Weast

Sr. Principal Engineer, Intel

VP Automated Vehicle Standards, Mobileye

MOBILEYE
An Intel Company

GOALS

An Automated Vehicle must be

- SAFE
- USEFUL
- TRANSPARENT
- SCALABLE

3 FACETS OF AUTOMATED DRIVING

SENSE

- Perception of the complete environment
- The raw material

PLAN

- Decision-making
- Analyze the raw material, and what action to take

ACT

- Execute the plan
- Control acceleration, braking, steering

SAFETY VALIDATION

How would you demonstrate that an automated vehicle is safe?

FUNCTIONAL SAFETY STANDARDS

System-level Safety

- ISO 26262 guides electric, electronic, and software quality
- Reduce chance of system faults, mitigate those that do occur
- Essential, but not the full picture

NORMATIVE SAFETY STANDARDS

Algorithm-level safety

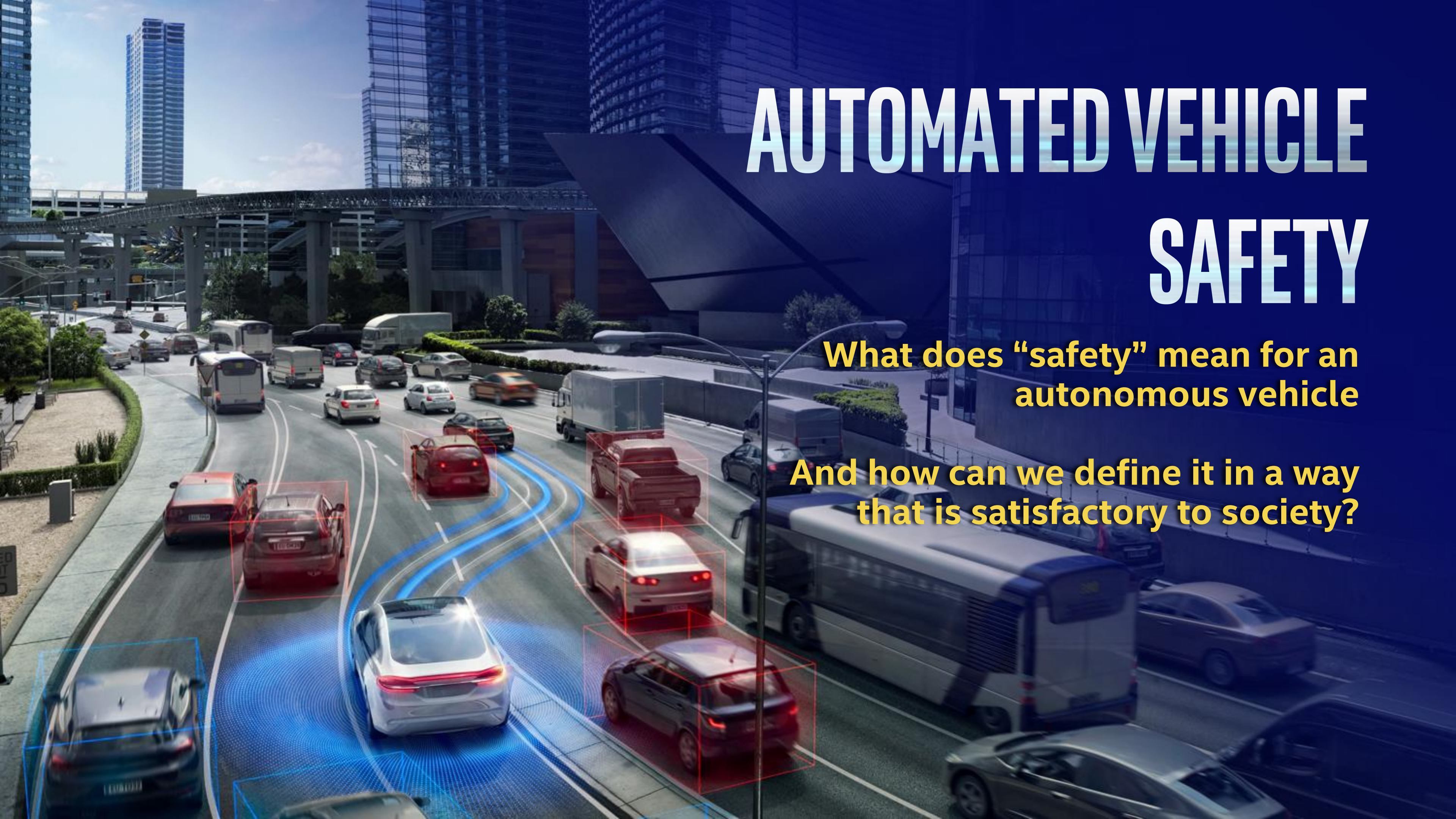
- Process to identify classes of safety violations not covered by ISO 26262
- Open to interpretation, which would result in different definitions of "safety"

SAFETY OF THE INTENDED FUNCTION (SOTIF)

AUTOMATED VEHICLE SAFETY

What does “safety” mean for an autonomous vehicle

And how can we define it in a way that is satisfactory to society?



HOW WOULD YOU DEFINE "SAFETY" FOR AN AV?

First try

Self-driving cars should be
statistically better than a human driver

THE STATISTICAL APPROACH TO SAFETY

The more miles I drive, the safer I am

Probability ρ of fatality / 1 hour of driving in U.S.	To demonstrate ρ an AV must drive	Averaging 30mph, that amounts to
10^{-6}	$\frac{1}{\rho}$ hours	~30m miles

To build trust,
we need to be better
by 2-3 orders of magnitude

1 Kalra, Nidhi and Susan M. Paddock, Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?. Santa Monica, CA: RAND Corporation, 2016. https://www.rand.org/pubs/research_reports/RR1478.html

THE STATISTICAL APPROACH TO SAFETY

The more miles I drive, the safer I am

For society to accept AVs, ρ should be	Averaging 30mph, that amounts to	100 cars driving 24/7/365 would take
10^{-9}	~30b miles	Over a **millennium**

Not Safe

Not just once:
Every update of hardware & software

Not Scalable

1 Kalra, Nidhi and Susan M. Paddock, Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?. Santa Monica, CA: RAND Corporation, 2016. https://www.rand.org/pubs/research_reports/RR1478.html

MILES DRIVEN

The more miles I drive without a crash, the safer I am

Miles driven here

Not the same as here

DISENGAGEMENTS

Minimize the number of times the ADS fails and requires a takeover

Why it's insufficient

- Similar to miles driven, depends on where & when
- Incentive to avoid the tough environments likely to trigger disengagements

HOW WOULD YOU DEFINE "SAFETY" FOR AN AV?

Second try

Develop other machine-friendly methods to define and prove safety

OTHER METHODS: SIMULATION

Why simulation alone cannot fully validate planning

- While sensing validation thrives in simulation, planning faces limitations
- Driving is a multi-agent system, to simulate it accurately is to simulate human behavior

WE CANNOT PROVABLY ACCURATELY SIMULATE THE REAL WORLD

OTHER METHODS: SCENARIOS

Expose the AV to the complete set of driving scenarios

Why it's insufficient

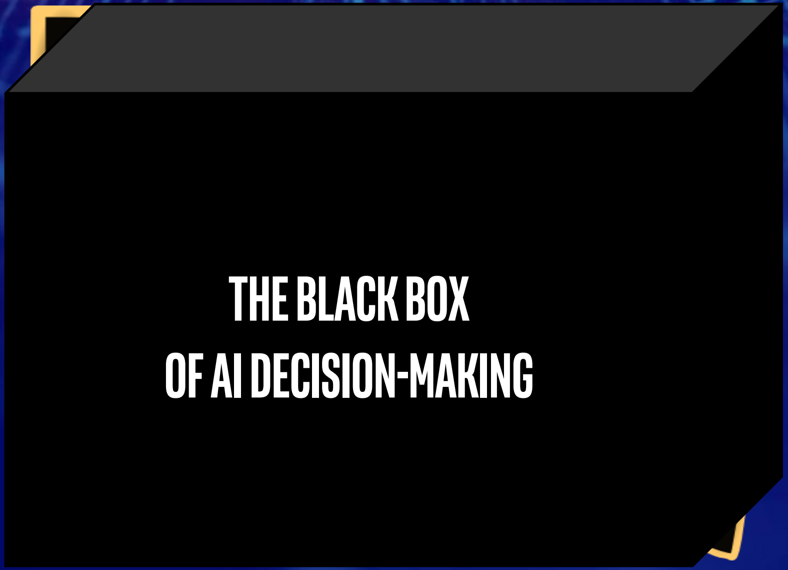
- Have to generalize; my list covers any other similar but omitted scenarios
- Difficult to draw the appropriate line between abstract & concrete scenarios
- Incents industry to build to the test

Pre vs. Post Deployment

- Pre-deployment testing assumes that it's possible to test everything
- And that nothing new will come up post-deployment

OTHER METHODS: PROPRIETARY

Trust me!



... PAY NO ATTENTION TO THAT MAN BEHIND THE CURTAIN

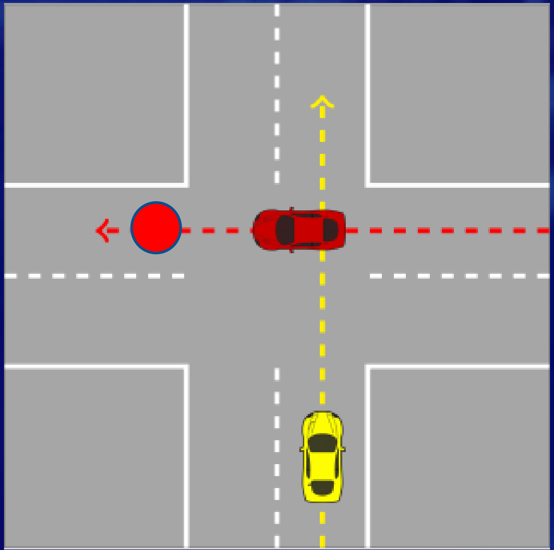
HOW WOULD YOU DEFINE "SAFETY" FOR AN AV?

Third try

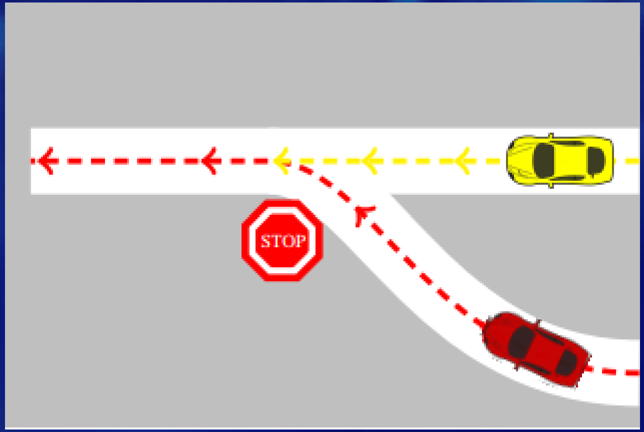
The AV only needs to strictly obey the rules of the road

SHOULD THE AV "FOLLOW THE RULES OF THE ROAD"?

- Traffic light



- Right of way



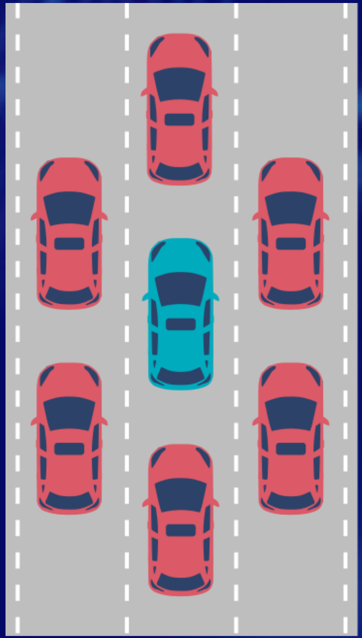
HOW WOULD YOU DEFINE "SAFETY" FOR AN AV?

Fourth try

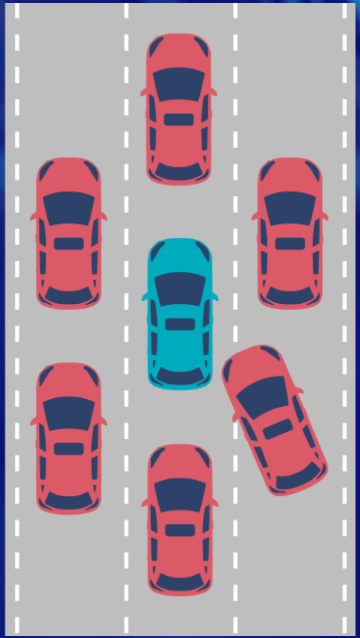
Avoid accidents **at all costs**

THE AV MUST AVOID ACCIDENTS AT ALL COSTS

Before



After



THE AV MUST AVOID ACCIDENTS AT ALL COSTS

https://www.youtube.com/watch?v=ctoBivu2NSE

THE AV MUST AVOID ACCIDENTS AT ALL COSTS

THE AV MUST AVOID ACCIDENTS AT ALL COSTS

WE NEED SOMETHING BETTER

And we're not the only ones who think so

ACADEMIA

"Specify unsafe regions for safety, specify safe regions for functionality. A 'safety envelope'"[1]

– Prof. Philip Koopman, CMU

THINK TANKS

"There is currently no accepted, industry-wide approach to [safety] demonstration"[2]

– Measuring Automated Vehicle Safety, RAND Corporation

GOVERNMENT

"The metrics that are most widely used by self-driving car developers -- miles driven and the frequency of human intervention -- alone are insufficient to demonstrate the safety of an autonomous automobile."[3]

- Derek Kan, Undersecretary of Transportation for Policy

1 Koopman, Philip. "Highly Autonomous Vehicle Validation: It's more than just road testing!" Carnegie Mellon University. Edge Case Research, LLC. 2017.
2. Fraade-Blanar, Laura, Marjory S. Blumenthal, James M. Anderson, and Nidhi Kalra, Measuring Automated Vehicle Safety: Forging a Framework. Santa Monica, CA: RAND Corporation, 2018. https://www.rand.org/pubs/research_reports/RR2662.html.
3 Beene, Ryan. "Self-driving Car Industry Needs Better Metrics, DOT Official Says." Bloomberg, October 23, 2018.

HOW DO HUMANS DO IT?

A HUMAN COMMON-SENSE DEFINITION OF DRIVING SAFELY

An AV should, at all times, drive carefully enough so it will never be the cause of an accident, and drive cautiously enough such that it should be able to compensate for reasonable mistakes of others.

RESPONSIBILITY SENSITIVE SAFETY (RSS)

An open and transparent industry standard that provides a verifiable safety check for AV decision-making

FORMALIZE

Human notions of safe driving

IDENTIFY

A Dangerous Situation

EXECUTE

The Appropriate Response

Keep a safe distance longitudinally & laterally

Safe distance compromised in both directions

Brake to restore safe longitudinal distance

RSS: A FORMAL MODEL FOR AV SAFETY

RSS is:

- A mathematical model that formalizes a “common sense” interpretation of safe driving
  - What is a Dangerous Situation?
  - What is the proper response to a Dangerous Situation?
  - What does it mean to be reasonably cautious?
  - What assumptions can the AV make about the behavior of others?

WHERE DOES RSS FIT?

SENSE

PLAN

- Analyze the raw material, and consider actions
- Make a Decision

RSS IS A CHECK FOR PLANNING SAFETY

ACT

- Execute the plan
- Control acceleration, braking, steering

PLANNING gets you from point A to point B

RSS helps keep you safe along the way

APPROACH TO VERIFICATION

RSS CAN BE USED IN ANY MECHANISM FOR VERIFICATION

BALANCING SAFETY AND USEFULNESS

THE BALANCING ACT BETWEEN SAFE & USEFUL

THE BALANCING ACT BETWEEN SAFE & USEFUL

We have a tight window, but we have a reasonable expectation that car behind us will adjust

Brakes to keep safe distance

Before continuing

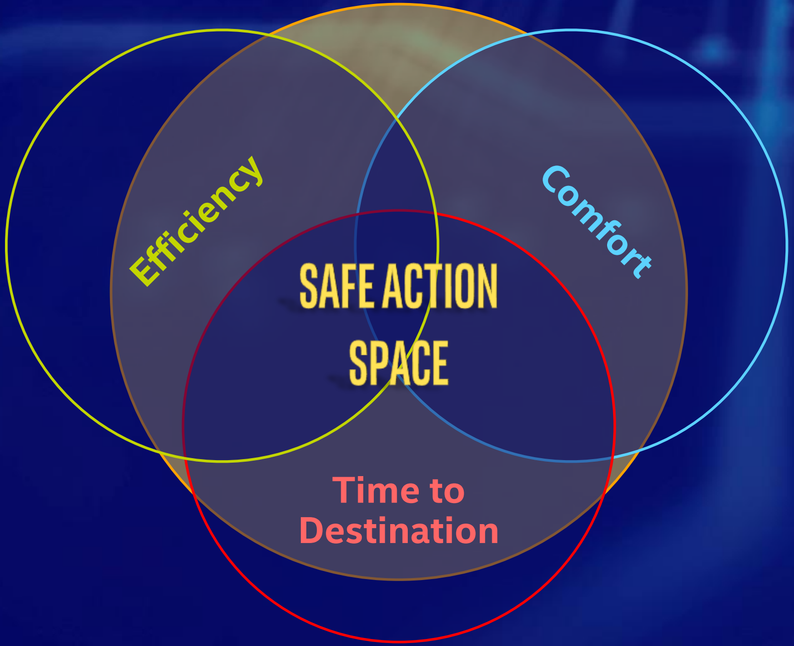
SAFE ACTION SPACE

How to maximize the safe actions available to the driving policy

- Safe action space: the set of all possible actions the AV can take that are safe
- Ideally: the AVs driving policy aligns and can propose any action within that space

SAFE ACTION SPACE

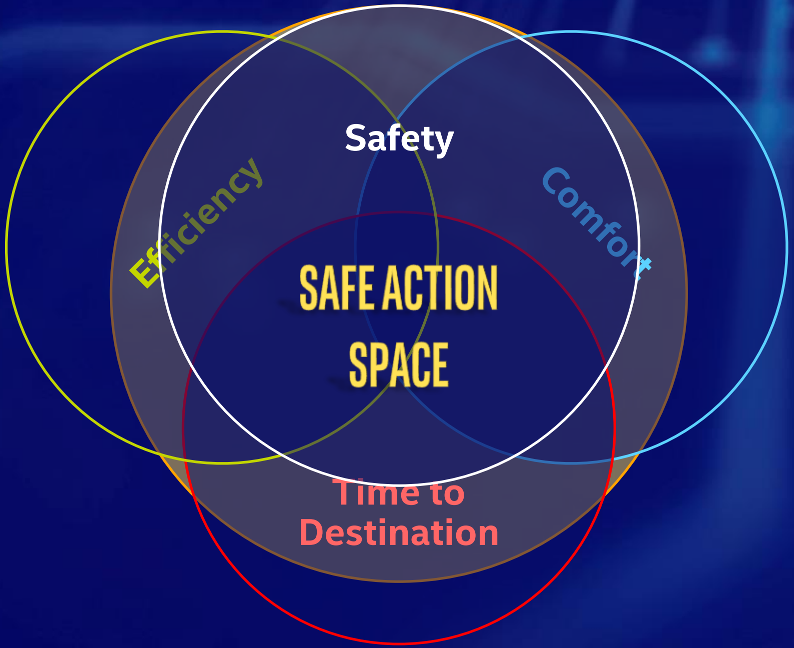
How do AVs today decide what actions to take?



- Driving policies learn with a Reward Function
- Motives/weights dictate what kind of driving experience the AV produces
- Without incorporating safety, some proposed actions will fall outside our safe action space

SAFE ACTION SPACE

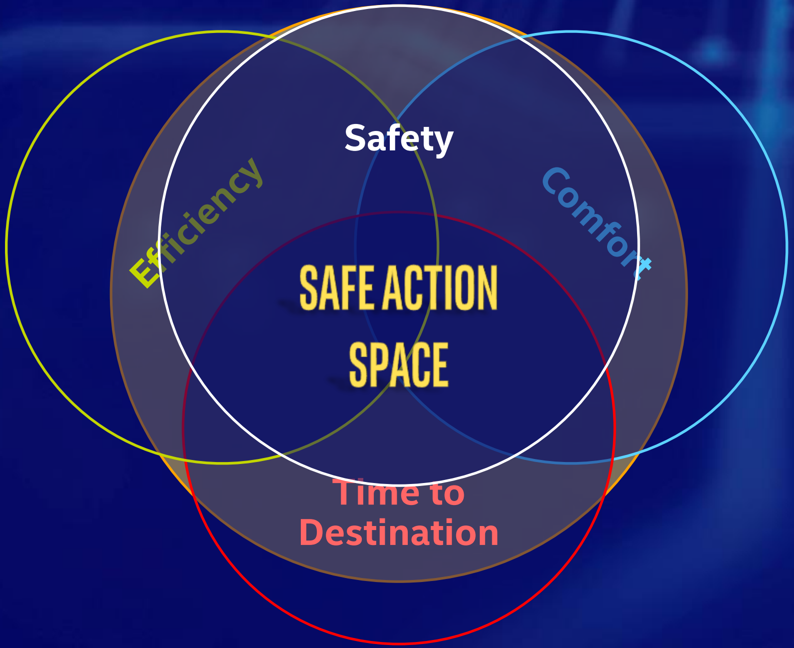
What if we add safety to the Reward Function?



- Adding safety to the Reward Function constrains the safe action space
- Safety now a competing interest in decision-making
- Now policy is overly-conservative, and still potentially unsafe

SAFE ACTION SPACE

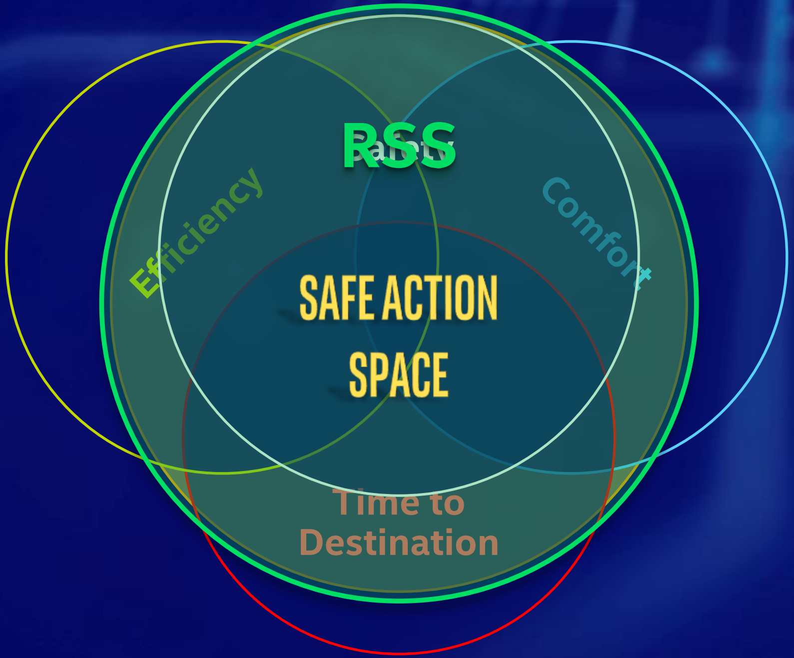
Safety cannot be left to proprietary chance



- How (or whether) an AV gets from point A to point B should be a proprietary differentiator
- Safety should be an open, transparent industry standard

SAFE ACTION SPACE

RSS is our missing layer



- Decouple safety from decision-making
- RSS becomes safety-check layer between driving policy and actuation
- RSS acts as the filter that defines safety

SAFE ACTION SPACE

RSS is our missing layer

- Decouple safety from decision-making
- RSS becomes a standard safety-check layer between proprietary driving policy and actuation
- RSS acts as the filter that defines safety for the industry

BASIC PRINCIPLES OF A SAFE AV

Rules we formalize in RSS

1. **Keep a safe distance from the car in front of you**
2. Leave time and space for others in lateral maneuvers
3. Exhibit caution in occluded areas
4. Right-of-Way is given, not taken
5. If you can safely avoid an accident without causing another you must do so

DEFINE SAFE LONGITUDINAL DISTANCE

$$\mathbf{d}_{\mathbf{min}} = \left[v_r \rho + \frac{1}{2} \alpha_{max} \rho^2 + \frac{(v_r + \rho \alpha_{max})^2}{2\beta_{min}} - \frac{v_f^2}{2\beta_{max}} \right]_+$$

DEFINE SAFE LONGITUDINAL DISTANCE

$$d_{min} = \left[v_r \rho + \frac{1}{2} \alpha_{max} \rho^2 + \frac{(v_r + \rho \alpha_{max})^2}{2\beta_{min}} - \frac{v_f^2}{2\beta_{max}} \right]_+$$

v_r Rear car (c_r) velocity

v_f Front car (c_f) velocity

DEFINE SAFE LONGITUDINAL DISTANCE

$$d_{min} = \left[v_r \rho + \frac{1}{2} \alpha_{max} \rho^2 + \frac{(v_r + \rho \alpha_{max})^2}{2\beta_{min}} - \frac{v_f^2}{2\beta_{max}} \right]_+$$

ρ Vehicle response time

β_{min} Min braking for c_r to apply to avoid colliding with c_f

DEFINE SAFE LONGITUDINAL DISTANCE

$$d_{min} = \left[v_r \rho + \frac{1}{2} \alpha_{max} \rho^2 + \frac{(v_r + \rho \alpha_{max})^2}{2\beta_{min}} - \frac{v_f^2}{2\beta_{max}} \right]_+$$

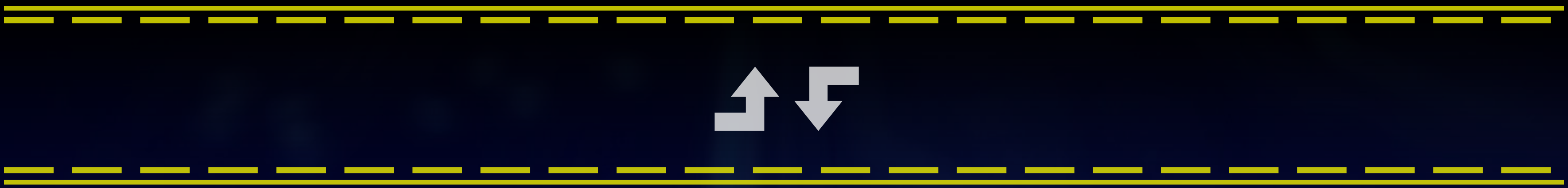
α_{max} Max acceleration during response time (for c_r)

β_{max} Max braking applied by c_f

not physical limits, but upper bounds on reasonable behavior

SAFE LONGITUDINAL DISTANCE (OPPOSITE DIRECTIONS)

$$d_{min} = \left(\frac{v_1 + v_{1,\rho}}{2}\right)\rho + \frac{v_{1,\rho}^2}{2\beta_{1,min}} + \left(\frac{|v_2| + v_{2,\rho}}{2}\right)\rho + \frac{v_{2,\rho}^2}{2\beta_{2,min}}$$

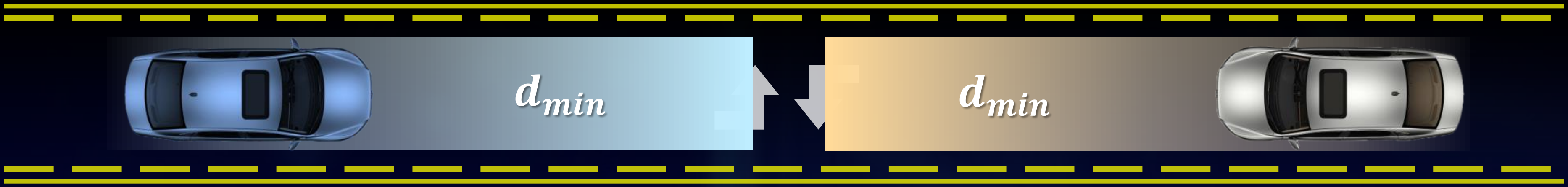


c_1 traveling with velocity v_1, $v_1 \geq 0$

c_2 traveling with velocity v_2, $v_2 < 0$

SAFE LONGITUDINAL DISTANCE (OPPOSITE DIRECTIONS)

$$d_{min} = \left(\frac{v_1 + \boldsymbol{v_{1,\rho}}}{2}\right)\rho + \frac{\boldsymbol{v_{1,\rho}^2}}{2\beta_{1,min}} + \left(\frac{|v_2| + \boldsymbol{v_{2,\rho}}}{2}\right)\rho + \frac{\boldsymbol{v_{2,\rho}^2}}{2\beta_{2,min}}$$



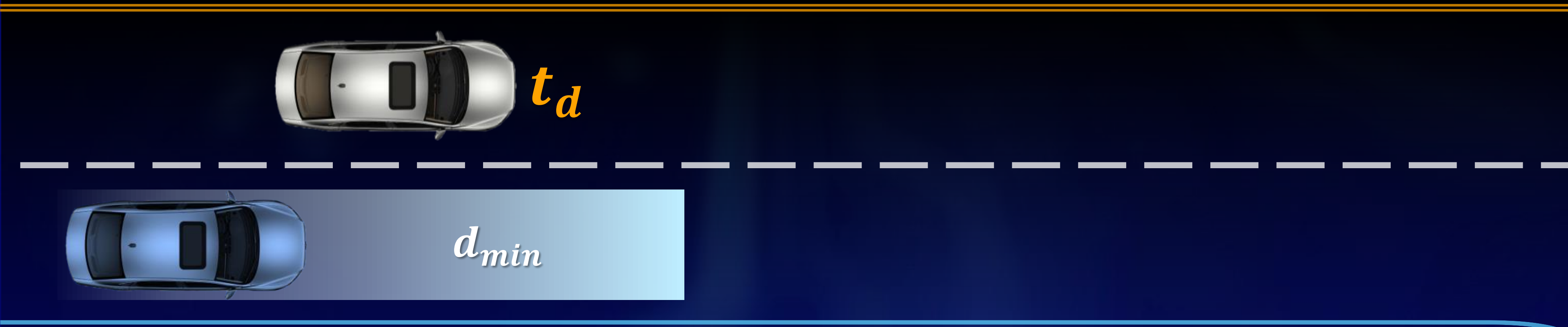
$$\boldsymbol{v_{1,\rho}} = v_1 + \rho\alpha_{max}$$

$$\boldsymbol{v_{2,\rho}} = |v_2| + \rho\alpha_{max}$$

Change in velocity during response time ρ

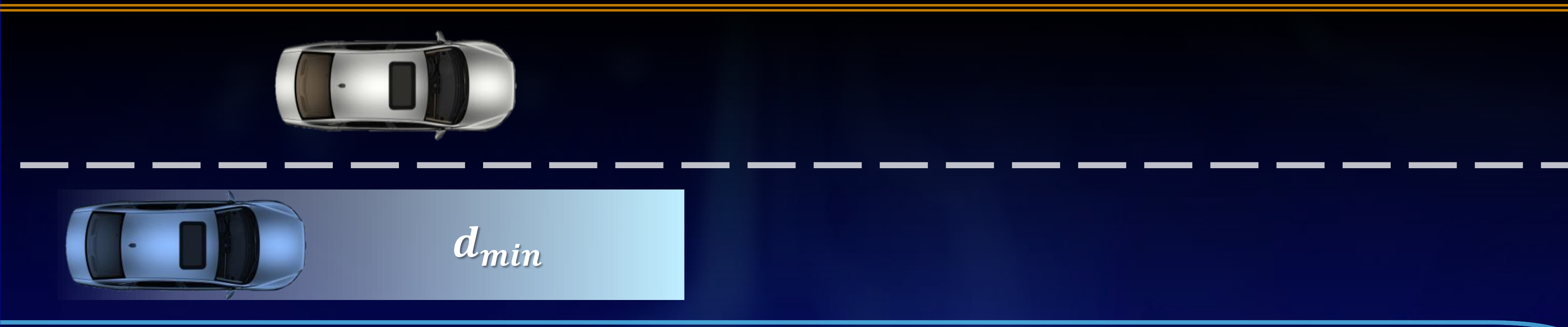
PROPER RESPONSE – LONGITUDINAL DANGER

The silver car has reached the Danger Threshold
(t_d is the last safe time before we enter a dangerous situation)



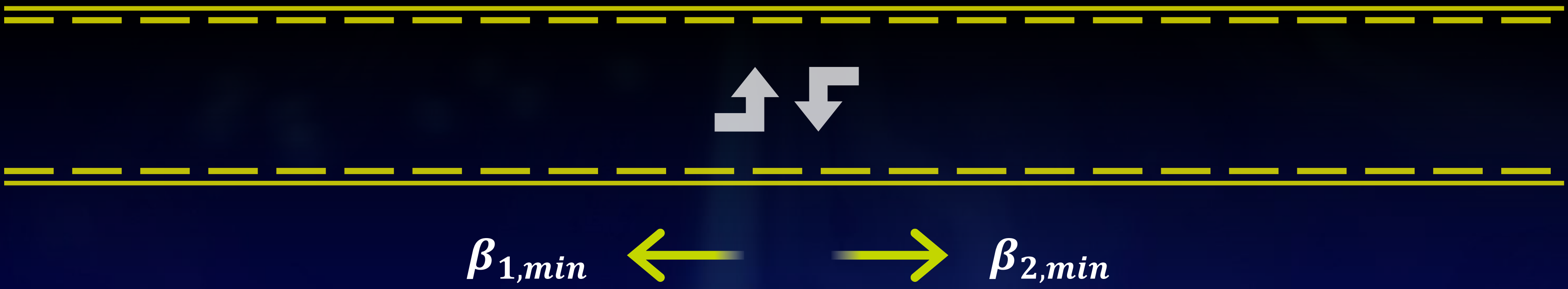
PROPER RESPONSE – LONGITUDINAL DANGER

Though the silver car initiated the dangerous situation,
the blue car still ought to brake to return to a safe distance



PROPER RESPONSE - OPPOSITE DIRECTION

If traveling in opposite directions,
both cars must apply the brakes to a full stop



BASIC PRINCIPLES OF A SAFE AV

Rules we formalize in RSS

1. Keep a safe distance from the car in front of you
2. **Leave time and space for others in lateral maneuvers**
3. Exhibit caution in occluded areas
4. Right-of-Way is given, not taken
5. If you can safely avoid an accident without causing another you must do so

DEFINE SAFE LATERAL DISTANCE

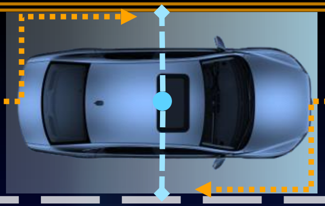
$$d_{min} = \mu + \left[\left(\frac{v_1 + v_{1,\rho}}{2} \right) \rho + \frac{v_{1,\rho}^2}{2\beta_{1,lat,min}} - \left(\left(\frac{v_2 + v_{2,\rho}}{2} \right) \rho + \frac{v_{2,\rho}^2}{2\beta_{2,lat,min}} \right) \right]$$

Cars usually perform small lateral movements,
Driving perfectly straight is impossible

DEFINE SAFE LATERAL DISTANCE

$$d_{min} = \mu + \left[\left(\frac{v_1 + v_{1,\rho}}{2} \right) \rho + \frac{v_{1,\rho}^2}{2\beta_{1,lat,min}} - \left(\left(\frac{v_2 + v_{2,\rho}}{2} \right) \rho + \frac{v_{2,\rho}^2}{2\beta_{2,lat,min}} \right) \right]$$

$l + \mu/2$



$l - \mu/2$

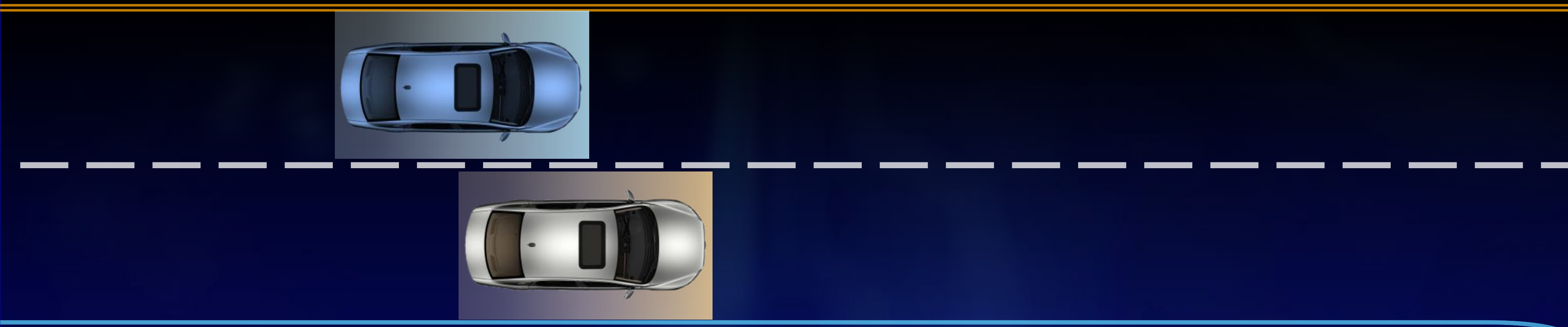
Given car's lateral position, ***l*** **is the lateral location at time** ***t***

μ represents our current lateral velocity

PROPER RESPONSE – LATERAL DANGER

If $t \in [t_d, t_d + \rho)$

Both cars must limit lateral acceleration

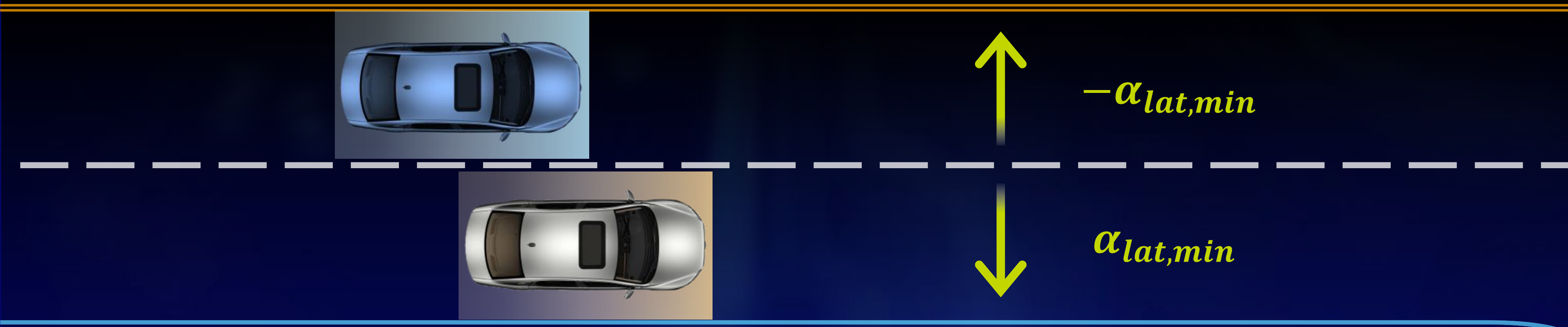


$$|\alpha| \leq \alpha_{lat,max}$$

PROPER RESPONSE – LATERAL DANGER

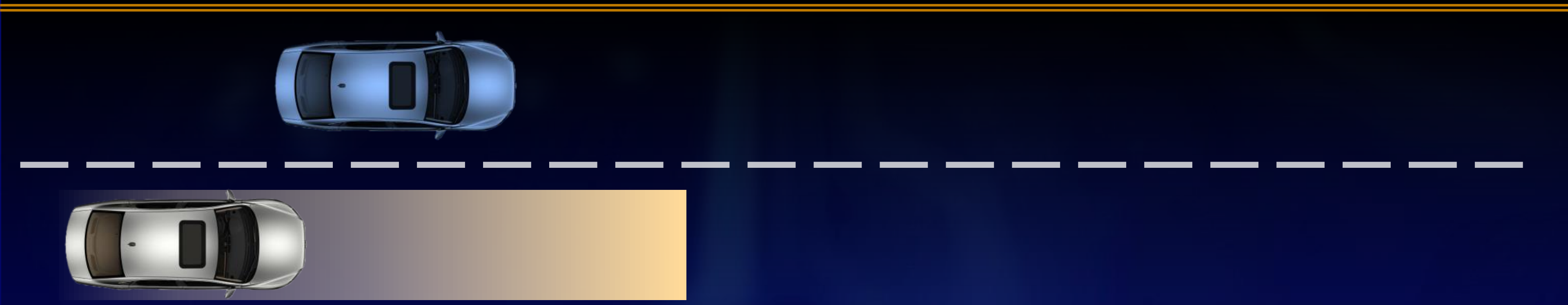
If $t \geq t_d + \rho$

Both cars must react



DEFINE DANGEROUS SITUATION

Time t is dangerous for cars c_1 , c_2 if *both* longitudinal and lateral distances between them are non safe

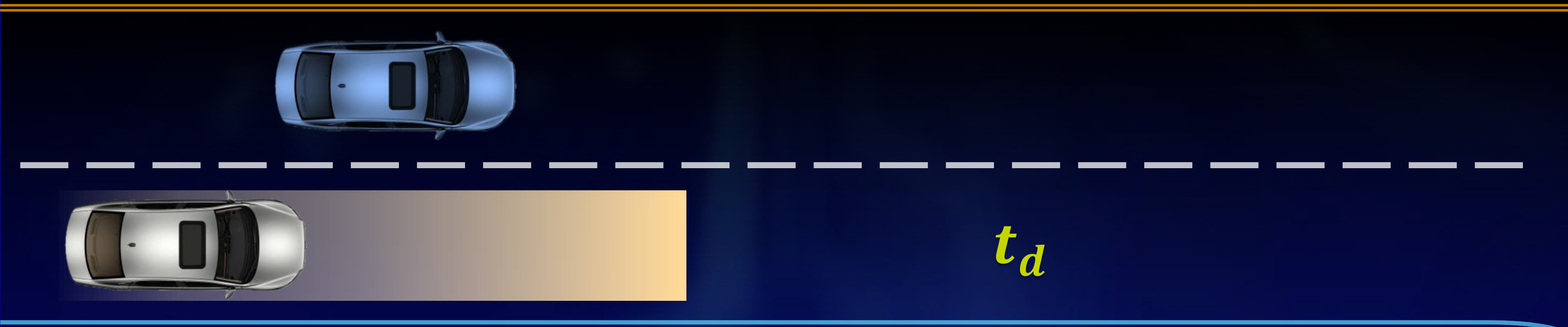


t is safe

t is dangerous

DEFINE DANGER THRESHOLD

Given a dangerous time t, its Danger Threshold, t_d, is the earliest non-dangerous time such that all times in the interval $(t_d, t]$ are dangerous



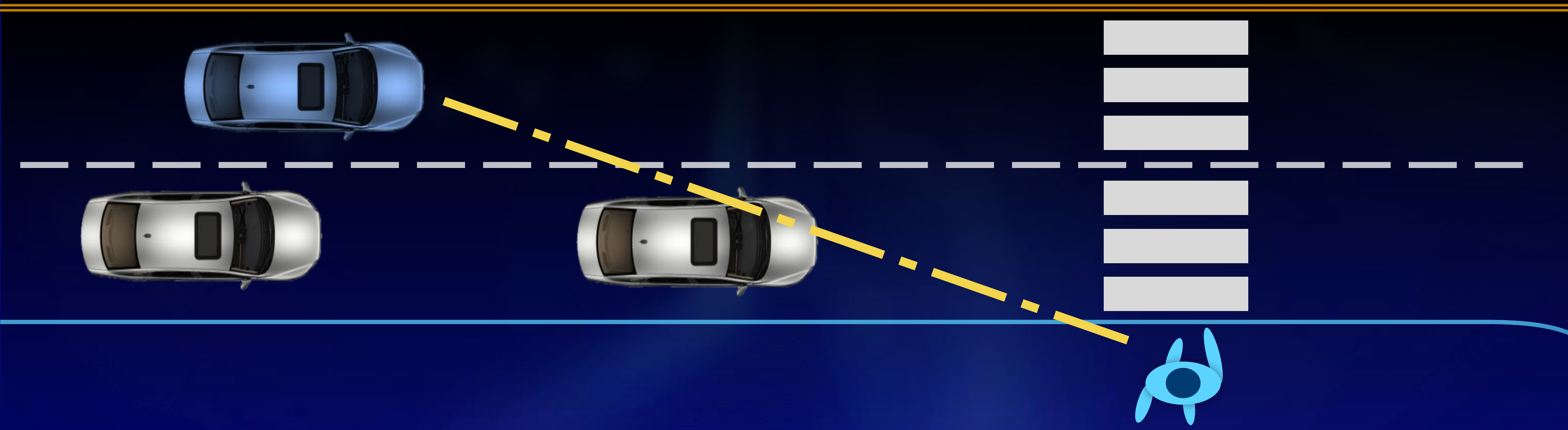
BASIC PRINCIPLES OF A SAFE AV

Rules we formalize in RSS

1. Keep a safe distance from the car in front of you
2. Leave time and space for others in lateral maneuvers
3. **Exhibit caution in occluded areas**
4. Right-of-Way is given, not taken
5. If you can safely avoid an accident without causing another you must do so

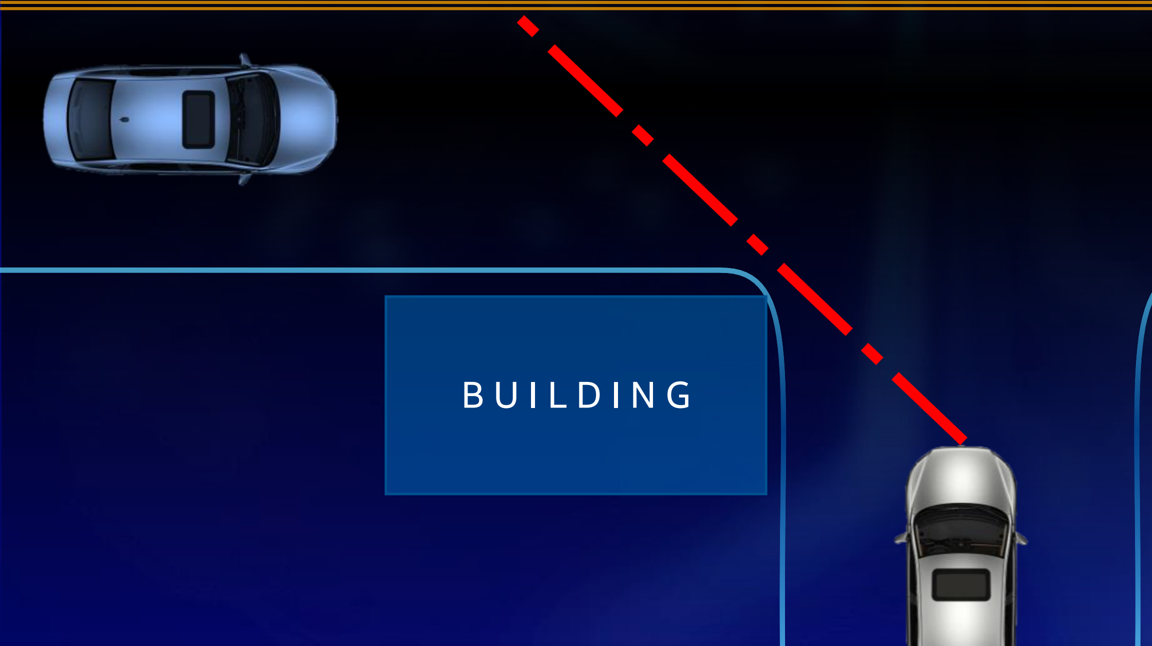
LIMITED VISIBILITY & OCCLUDED AREAS

When sensing capabilities are physically limited,
We must exhibit caution



LIMITED VISIBILITY – BLIND CORNER

Both cars assume a reasonable limit on the speed of the other



What is a reasonable assumption on the speed limit of the other?

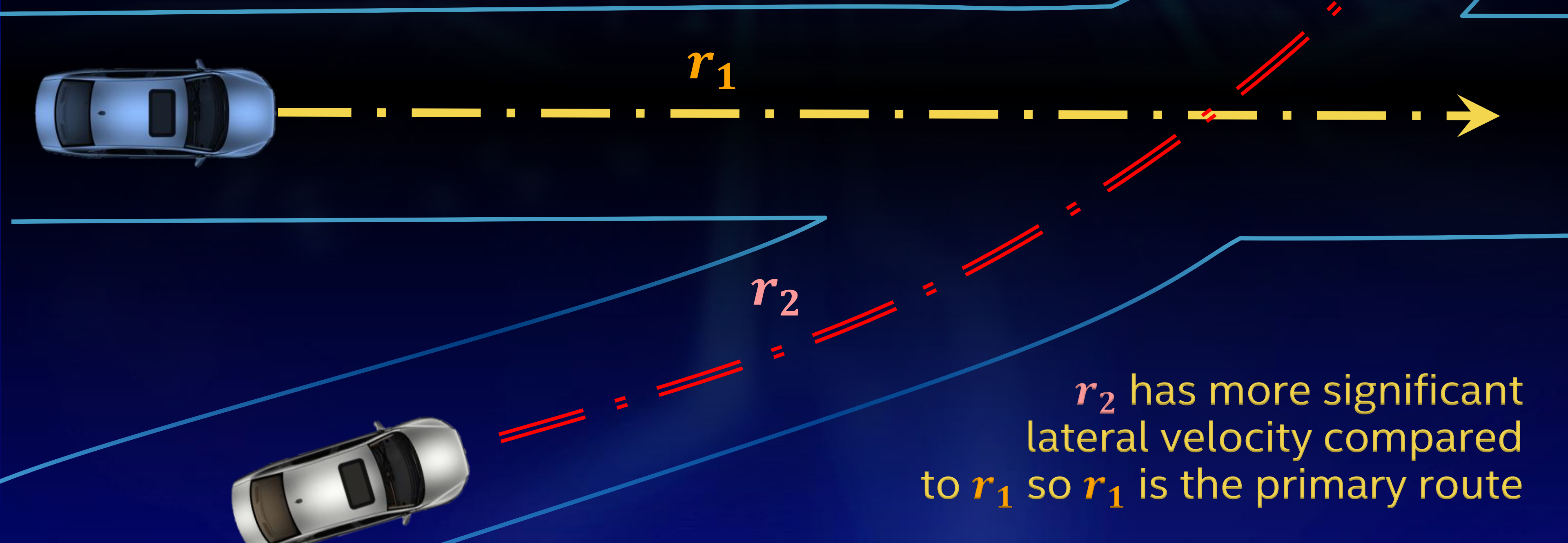
BASIC PRINCIPLES OF A SAFE AV

Rules we formalize in RSS

1. Keep a safe distance from the car in front of you
2. Leave time and space for others in lateral maneuvers
3. Exhibit caution in occluded areas
4. **Right-of-Way is given, not taken**
5. If you can safely avoid an accident without causing another you must do so

RIGHT – OF – WAY

How do we establish priority on roads with odd geometries?

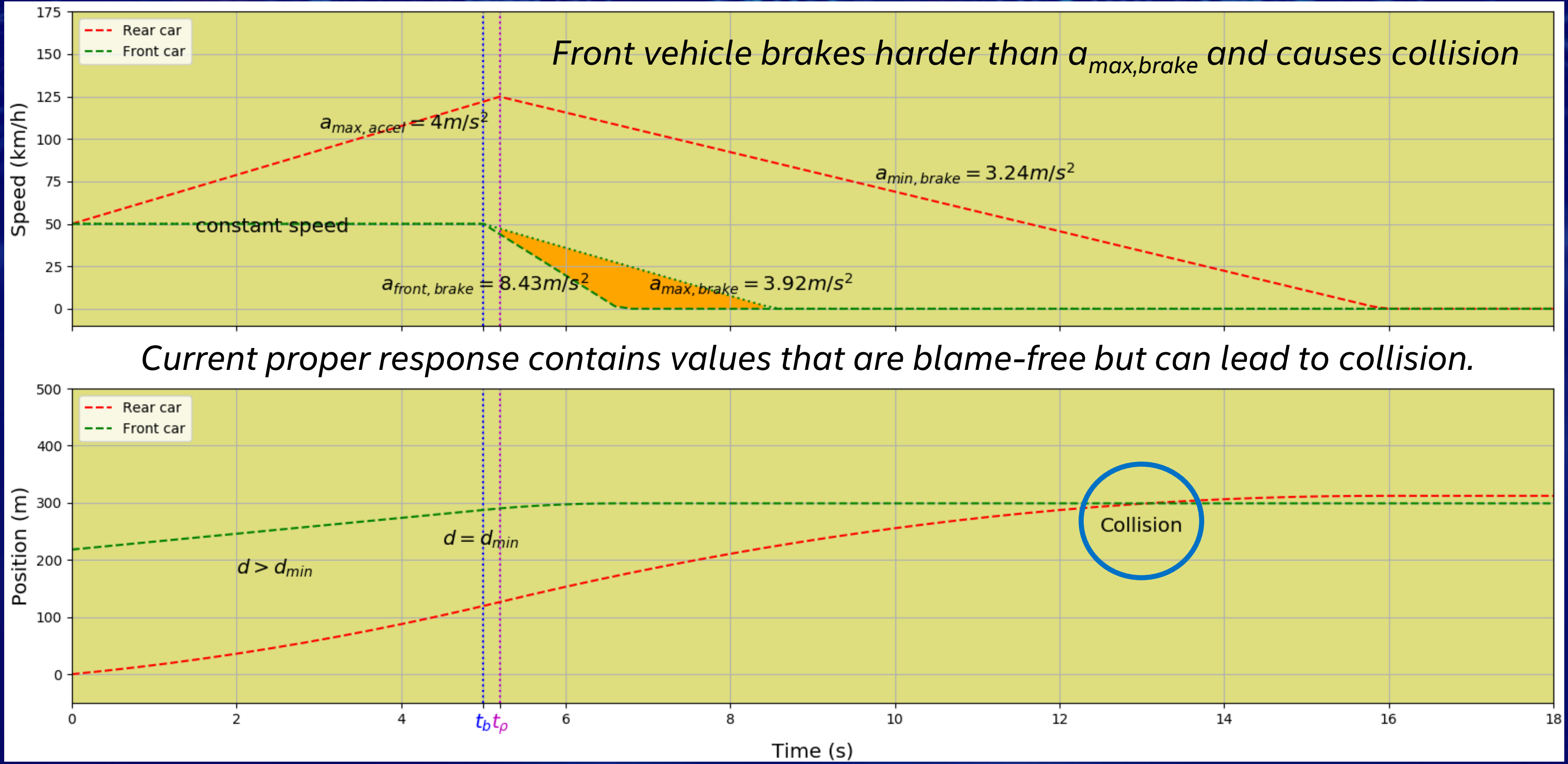


r_2 has more significant lateral velocity compared to r_1 so r_1 is the primary route

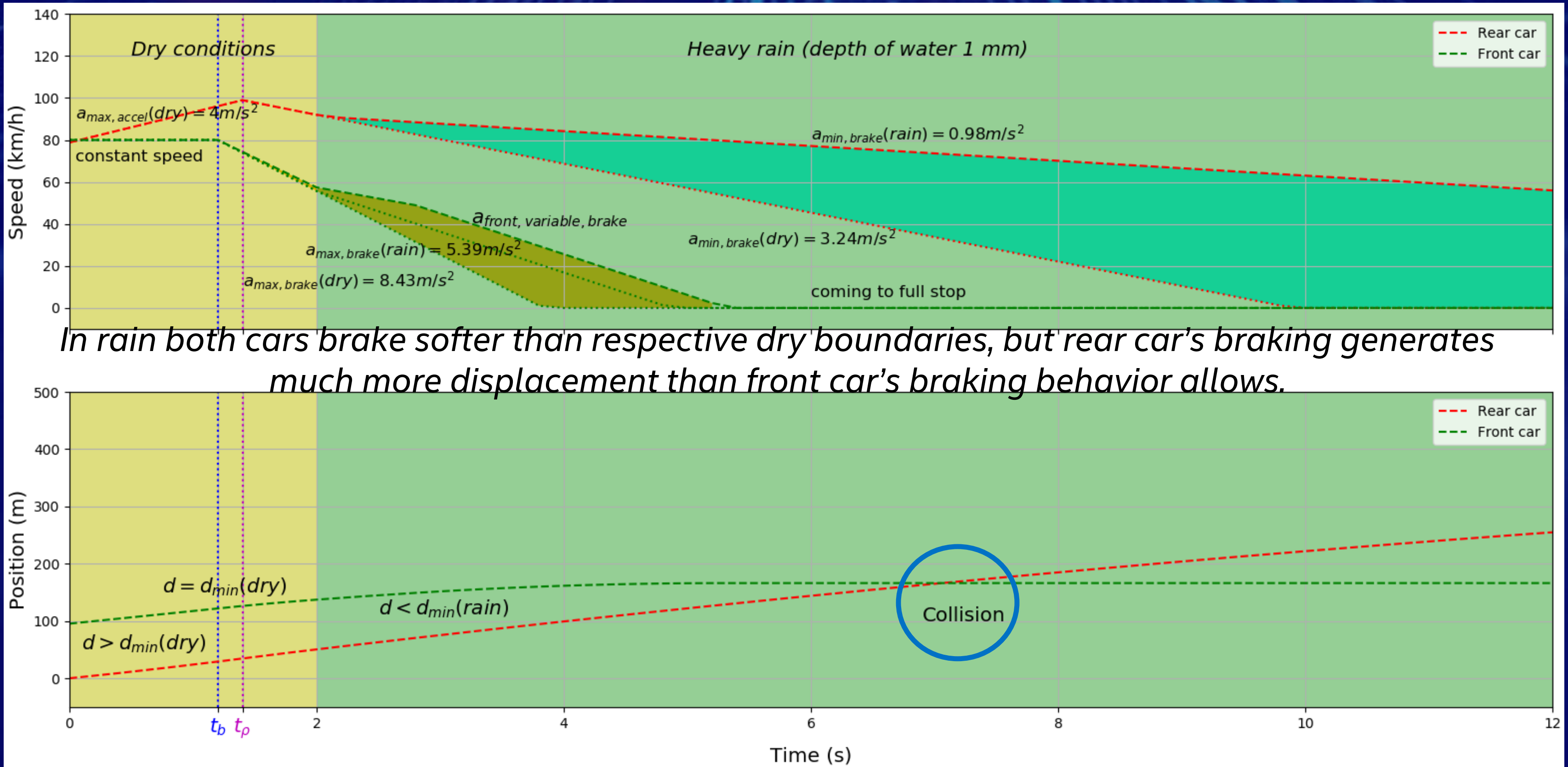
GRAY AREAS – WORK TO DO

Why this needs to be an open and transparent discussion

What if the front vehicle brakes >max,brake?



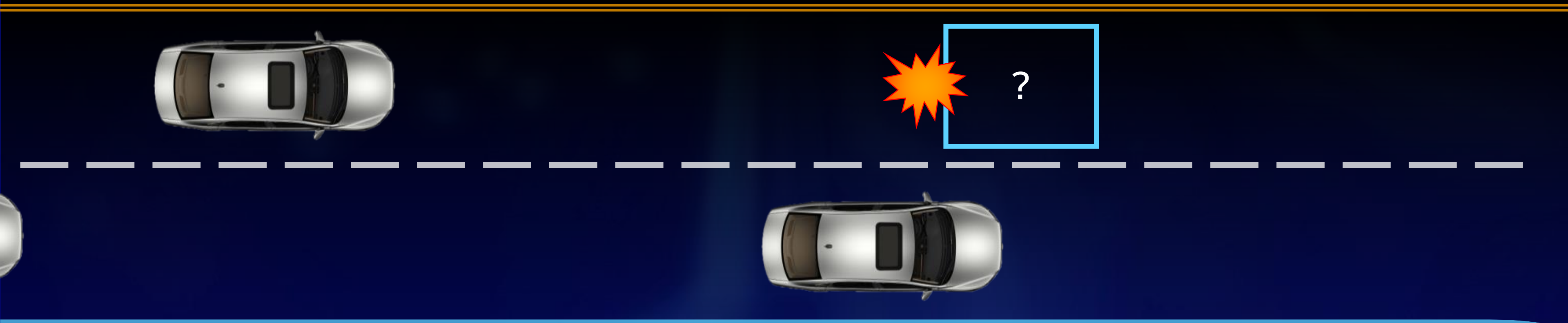
Discontinuities in Road Condition



In rain both cars brake softer than respective dry boundaries, but rear car's braking generates much more displacement than front car's braking behavior allows.

REASONABLE ASSUMPTIONS ON THE ROAD

Consider this:
An object on the road we only detect after its too late,
because the silver car changes lanes at the last moment



Should safe distance account for this worst-case scenario?

REASONABLE EXPECTATIONS ON THE ROAD

To keep a safe distance on a highway going ~65mph, a car would need more than 150 feet (~10 car lengths) to stop in time

Society would likely agree this is unreasonable… so what can the AV assume about others?

GRAY AREAS WITH PROPER RESPONSES

Proportional Responsibility
In some places, like the US, it is not always binary
We made the Proper Response, but are not “responsibility free”

Should safe distance account for the potential actions of the rear car?

AV SAFETY: AN ISSUE LARGER THAN ONE COMPANY

What are we doing

INDUSTRY

Engaging with customers, competitors and consortia to have an open dialogue on the safety assurance of AV's

GOVERNMENT / NGO'S

Understanding government and NHO expectations on transparency and measurable verification of AV's

ACADEMIA

RSS Research Centers at Universities in USA and PRC

REAL WORLD

Deploying RSS in our AV Fleet in some of the most challenging environments

FOR MORE INFORMATION

On a Formal Model of Safe and Scalable Self-driving Cars

Shai Shalev-Shwartz, Shaked Shammah, Amnon Shashua

Mobileye, 2017

Abstract

In recent years, car makers and tech companies have been racing towards self driving cars. It seems that the main parameter in this race is who will have the first car on the road. The goal of this paper is to add to the equation two additional crucial parameters. The first is standardization of safety assurance — what are the minimal requirements that every self-driving car must satisfy, and how can we verify these requirements. The second parameter is scalability — engineering solutions that lead to unleashed costs will not scale to millions of cars, which will push interest in this field into a niche academic corner, and drive the entire field into a "winter of autonomous driving". In the first part of the paper we propose a white-box, interpretable, mathematical model for safety assurance, which we call Responsibility-Sensitive Safety (RSS). In the second part we describe a design of a system that adheres to our safety assurance requirements and is scalable to millions of cars.

1 Introduction

The "Winter of AI" is commonly known as the decades long period of inactivity following the collapse of Artificial

RSS IN SUMMARY

SAFE TRANSPARENT AFFORDABLE USEFUL

An open and transparent industry standard that provides verifiable safety assurance for AV decision-making

- The industry must collaborate with governments and agree on what it means for an AV to drive safely
- RSS provides a starting point for a definition of what it means for an AV to drive safely
- RSS can be formally verified and so solves the statistical verification challenge with an open and measurable metric
- RSS is technology neutral compatible with any AV solution

Join us in this important effort to provide safety assurance for Automated Vehicles!

An Intel Company