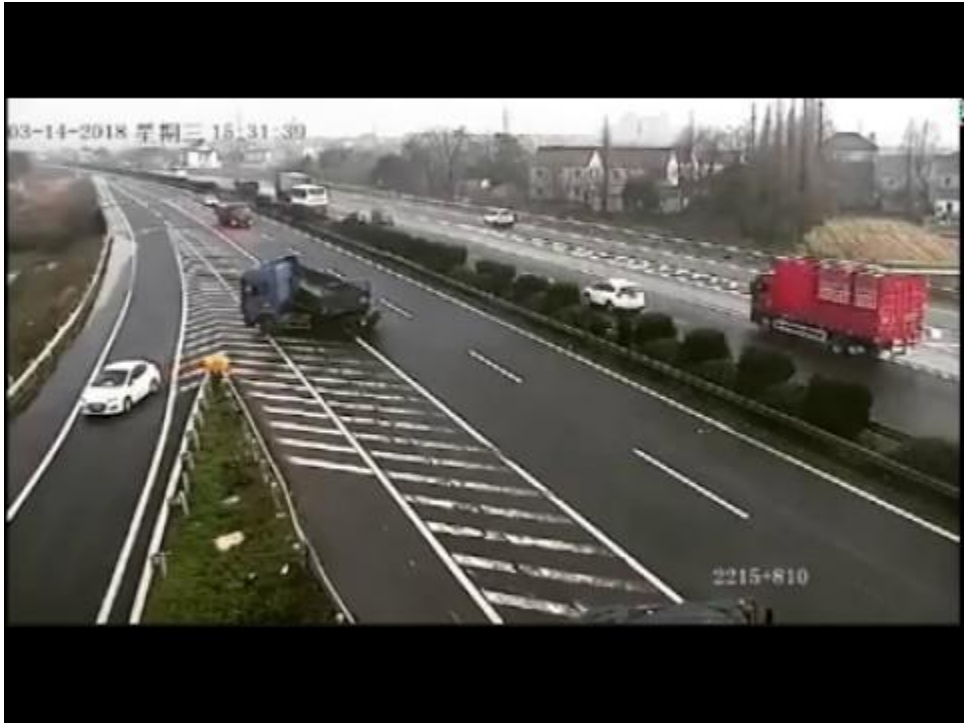
TOYOTA
RESEARCH INSTITUTE

Building an AV Safety Case

OSS 5, San Francisco, Feb. 28, 2019

Sagar Behere



https://www.reddit.com/r/nononono/comments/8ahc7r/running_late_to_work_cant_miss_my_exit/

Toyota Research Institute

We are showing what is possible when the limits to mobility are challenged...

...without claiming that anywhere/anytime autonomous driving just around the corner ;-)

TRI: Autonomy Capability

Guardian

A measure of how much the automated driving system helps to protect … while the human is driving.

Chauffeur

A measure of the degree to which the vehicle takes the primary responsibility for driving…

TOYOTA
RESEARCH INSTITUTE

Content of an AV Safety Case

Content of an AV Safety case

1. PHILOSOPHY
2. CONTEXT
3. DESIGN, IMPLEMENTATION
4. EVIDENCE
5. COVERAGE/RESIDUAL RISK
6. LARGER QUESTIONS

- Definition of safety
- Safety goals
- General approach to assurance

Content of an AV Safety case

1	PHILOSOPHY
2	CONTEXT
3	DESIGN, IMPLEMENTATION
4	EVIDENCE
5	COVERAGE/RESIDUAL RISK
6	LARGER QUESTIONS

- Operational Design Domain (ODD)
- Assumptions
- Operational procedures

Content of an AV Safety case

1	PHILOSOPHY
2	CONTEXT
3	DESIGN, IMPLEMENTATION
4	EVIDENCE
5	COVERAGE/RESIDUAL RISK
6	LARGER QUESTIONS

- What constitutes a safe design?
- What constitutes a safe implementation?
- What constitutes a safe development process?
- What properties must an AV possess in order to be considered safe?

Content of an AV Safety case

1 PHILOSOPHY

2 CONTEXT

3 DESIGN, IMPLEMENTATION

4 EVIDENCE

5 COVERAGE/RESIDUAL RISK

6 LARGER QUESTIONS

- Basis for evaluating a claim of safety
- Methods of evidence

Content of an AV Safety case

1 PHILOSOPHY

2 CONTEXT

3 DESIGN, IMPLEMENTATION

4 EVIDENCE

5 COVERAGE/RESIDUAL RISK

6 LARGER QUESTIONS

- Adequacy of safety properties in stated context
- Probability of safety violation

Content of an AV Safety case

1	PHILOSOPHY
2	CONTEXT
3	DESIGN, IMPLEMENTATION
4	EVIDENCE
5	COVERAGE/RESIDUAL RISK
6	LARGER QUESTIONS

- How safe is safe enough?
- Data sharing?
- Comparisons to human drivers?
- Cooperation and standardization?

Content of an AV Safety case

1	PHILOSOPHY
2	CONTEXT
3	DESIGN, IMPLEMENTATION
4	EVIDENCE
5	COVERAGE/RESIDUAL RISK
6	LARGER QUESTIONS

A credible AV safety case must provide rational evidence-based argumentation for each area

TOYOTA
RESEARCH INSTITUTE

Safety Philosophy

Quiz time: What is AV safety?

Quiz time: What is AV safety?

- What is the relationship between AV Safety and collisions?

  a. Does the presence of collision imply absence of safety?
  b. Does the absence of collision imply presence of safety?
  c. All of the above?
  d. None of the above?

- Don't leave the road; Don't hit things; Don't get hit ← Sufficient?

An example formulation

Within its ODD, ______________________ not outside of it

An example formulation

Within its ODD, ——— not outside of it

an AV shall not cause ——— be the primary cause of; do its best to avoid?

An example formulation

Within its ODD,

not outside of it

an AV shall not cause

be the primary cause of; do its best to avoid?

a foreseeable

what constitutes foreseeable?

An example formulation

Within its ODD,

not outside of it

an AV shall not cause

be the primary cause of; do its best to avoid?

a foreseeable

what constitutes foreseeable?

and

preventable

what constitutes preventable?

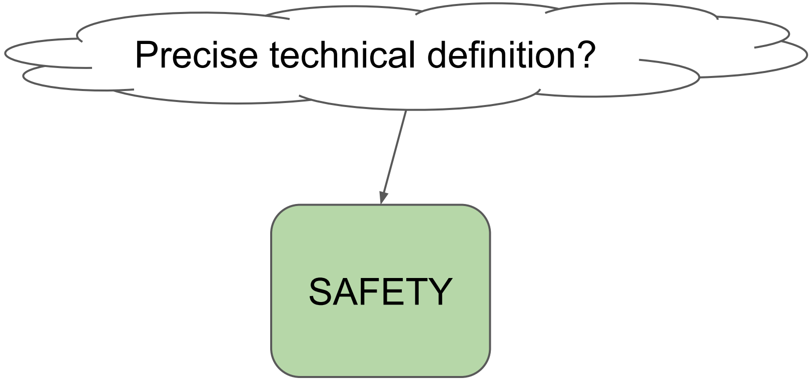
An example formulation

Within its ODD,	not outside of it
an AV shall not cause	be the primary cause of; do its best to avoid?
a foreseeable	what constitutes foreseeable?
and	
preventable	what constitutes preventable?
fatal incident.	why restrict to fatal?

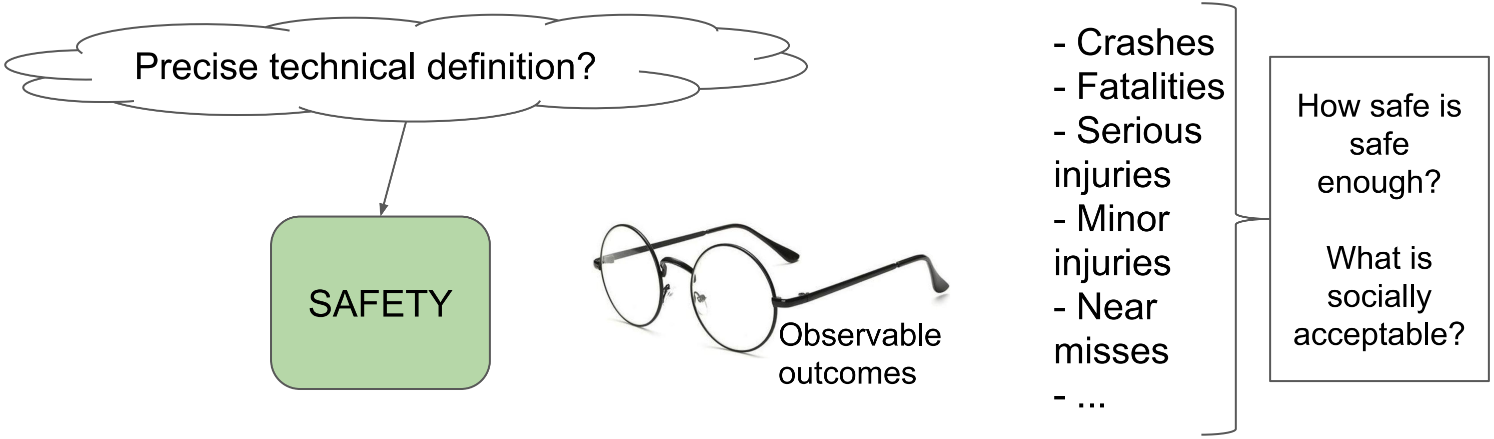
Identifying properties of a safe system

SAFETY

Identifying properties of a safe system



Identifying properties of a safe system



Precise technical definition?

SAFETY

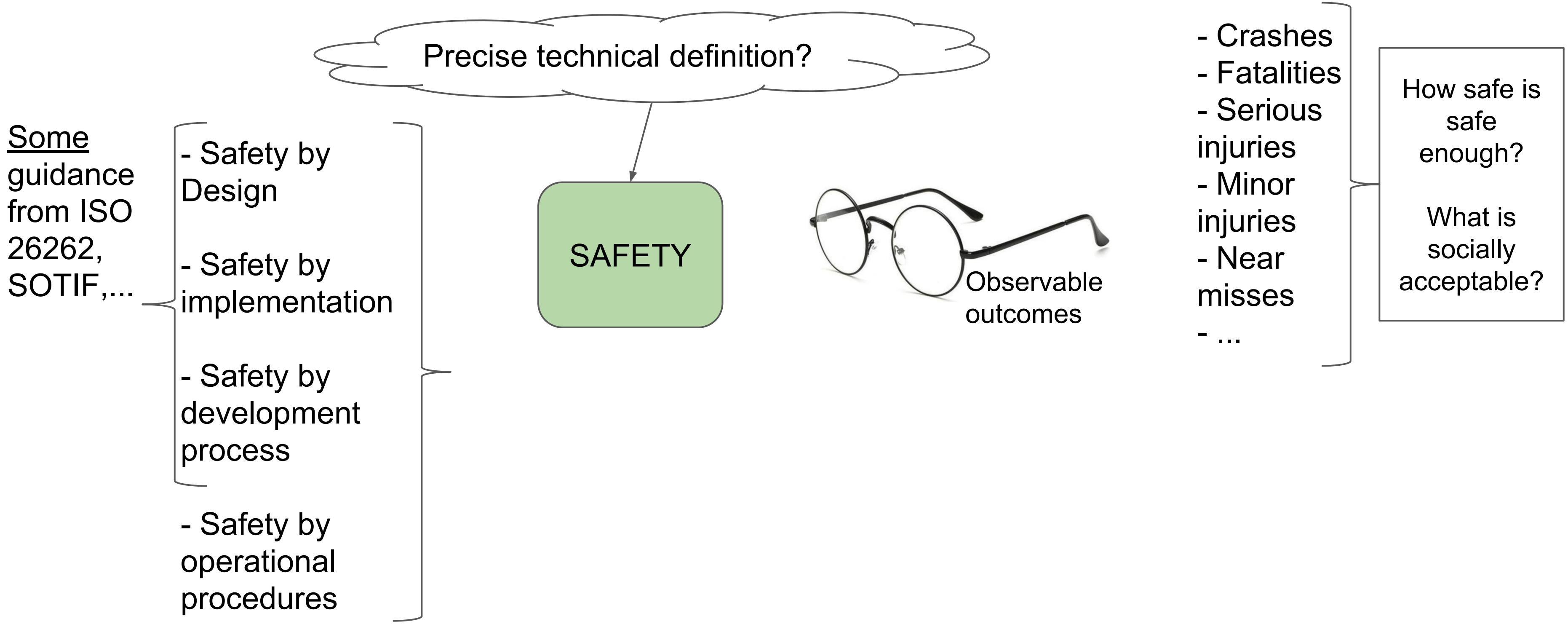
Observable outcomes

- Crashes
- Fatalities
- Serious injuries
- Minor injuries
- Near misses
- ...

How safe is safe enough?

What is socially acceptable?

Identifying properties of a safe system



Some guidance from ISO 26262, SOTIF,...

- Safety by Design
- Safety by implementation
- Safety by development process
- Safety by operational procedures

Precise technical definition?

SAFETY

Observable outcomes

- Crashes
- Fatalities
- Serious injuries
- Minor injuries
- Near misses
- ...

How safe is safe enough?

What is socially acceptable?

Identifying properties of a safe system

Some guidance from ISO 26262, SOTIF,...

- Safety by Design
- Safety by implementation
- Safety by development process
- Safety by operational procedures

Precise technical definition?

SAFETY

Observable outcomes

- Crashes
- Fatalities
- Serious injuries
- Minor injuries
- Near misses
- ...

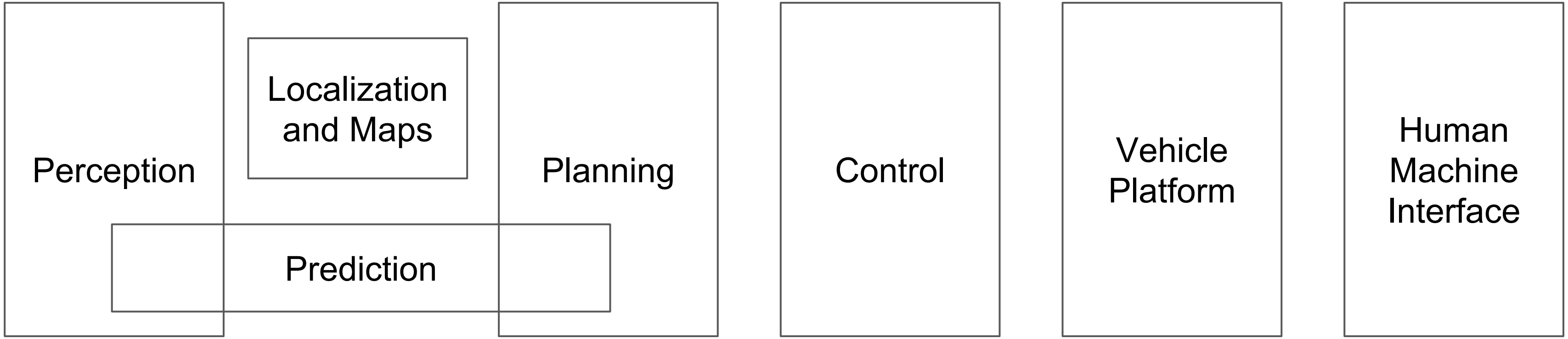
How safe is safe enough?

What is socially acceptable?

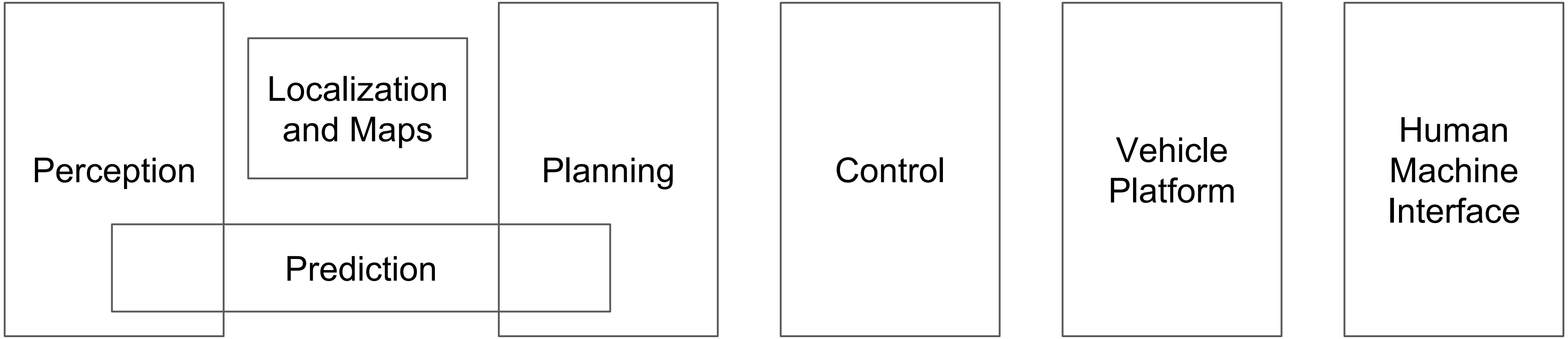
Is there a finite set of properties an AV can possess which eliminate/minimize undesired outcomes?

Design, Implementation

Core elements of AV architecture



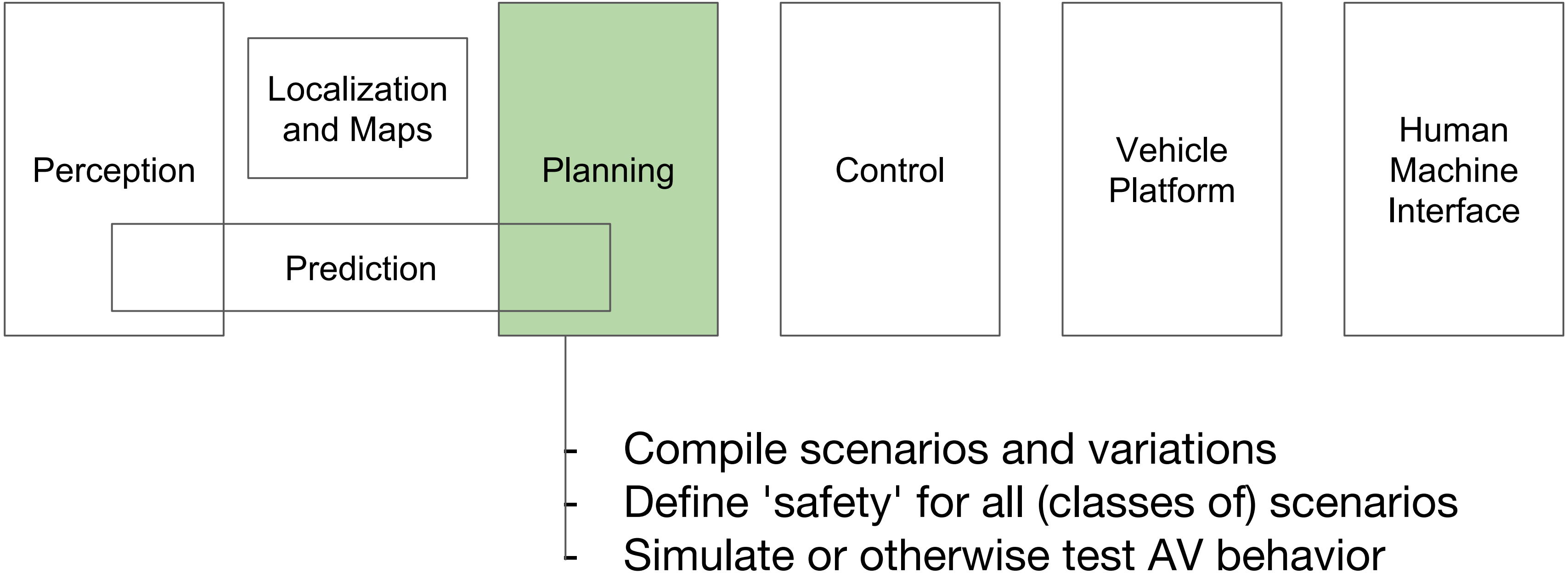
Core elements of AV architecture



Must reason deeply about needed safety of these, individually and collectively … in terms of design, implementation, and development process.

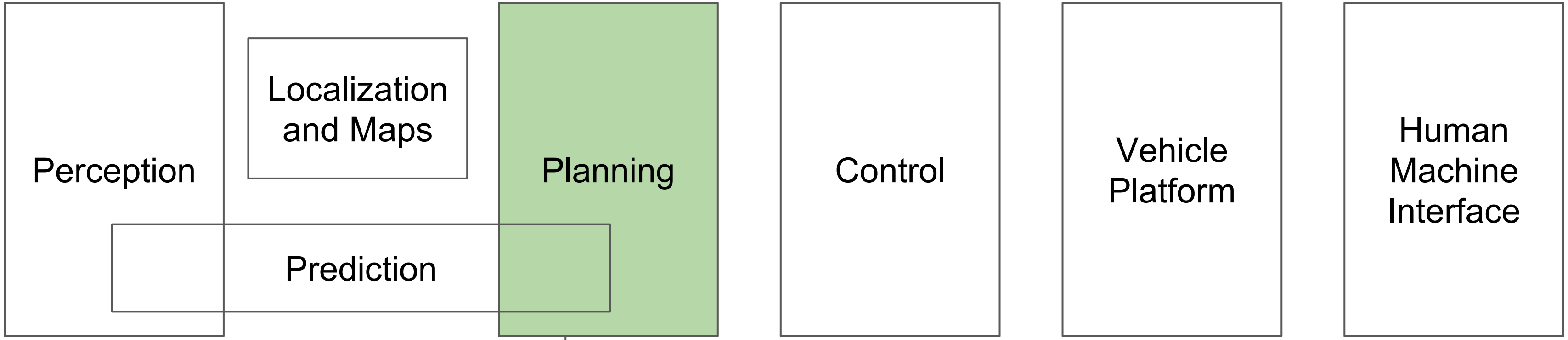
THINK: What would a handful of closed course tests show?

Example: Planning



- Compile scenarios and variations
- Define 'safety' for all (classes of) scenarios
- Simulate or otherwise test AV behavior

Example: Planning



Perception

Localization and Maps

Prediction

Planning

Control

Vehicle Platform

Human Machine Interface

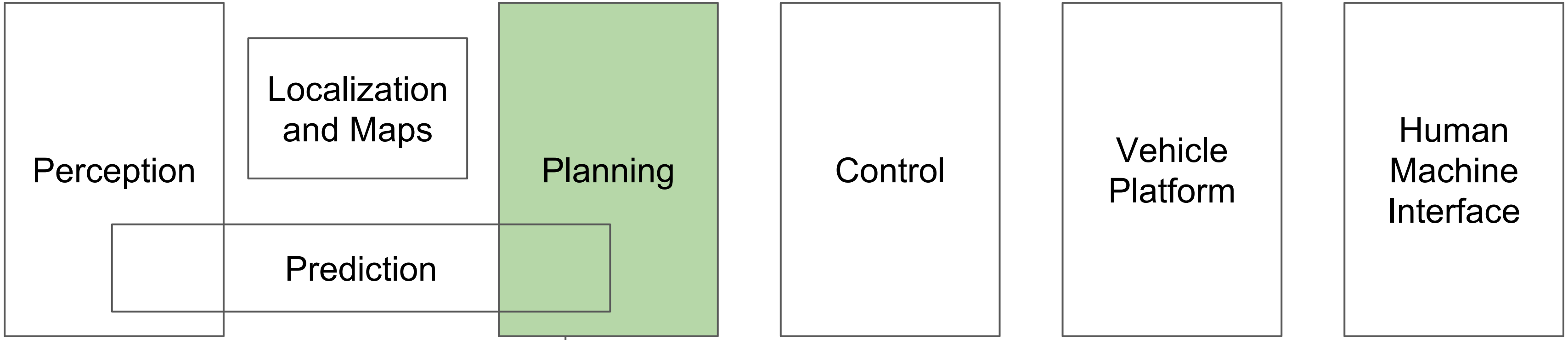
An NP-hard problem?

You can check a system solution fast enough, but can you find a solution that passes ALL current and future scenarios?

- Compile scenarios and variations
- Define 'safety' for all (classes of) scenarios
- Simulate or otherwise test AV behavior

Mathematically, this problem is intractable!
(Pragmatically, it is still useful)

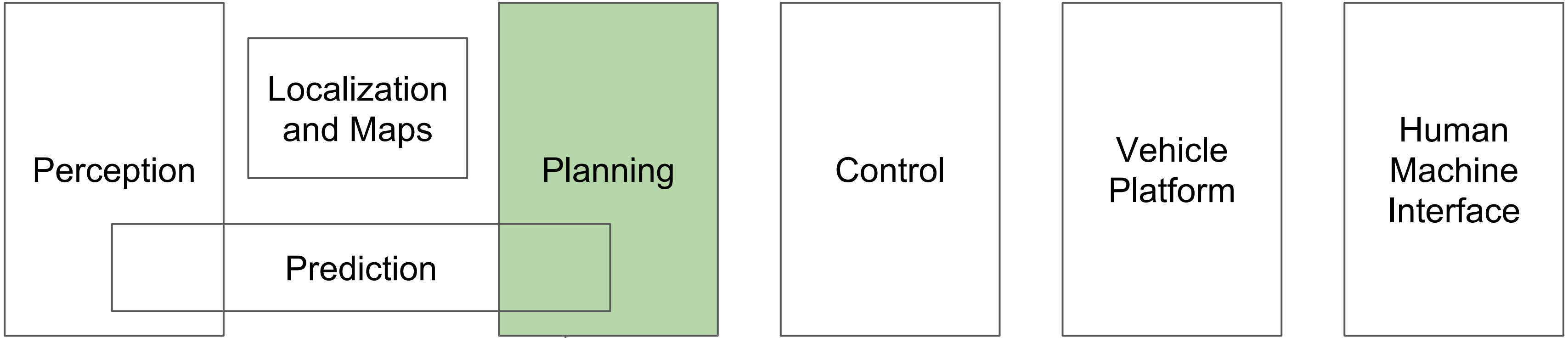
Making the problem tractable



- Find a finite set of planning rules
- Adherence to rules should avoid fatal incidents
- Prove that AV system will not violate rules

Tractable: ∞ possible accidents avoided by finite rule set

Making the problem tractable



Remember!

This still applies only to Planning, and assumes perfect inputs

⇨

- Find a finite set of planning rules
- Adherence to rules shall avoid fatal incidents
- Prove that AV system will not violate rules

Tractable: ∞ possible accidents avoided by finite rule set

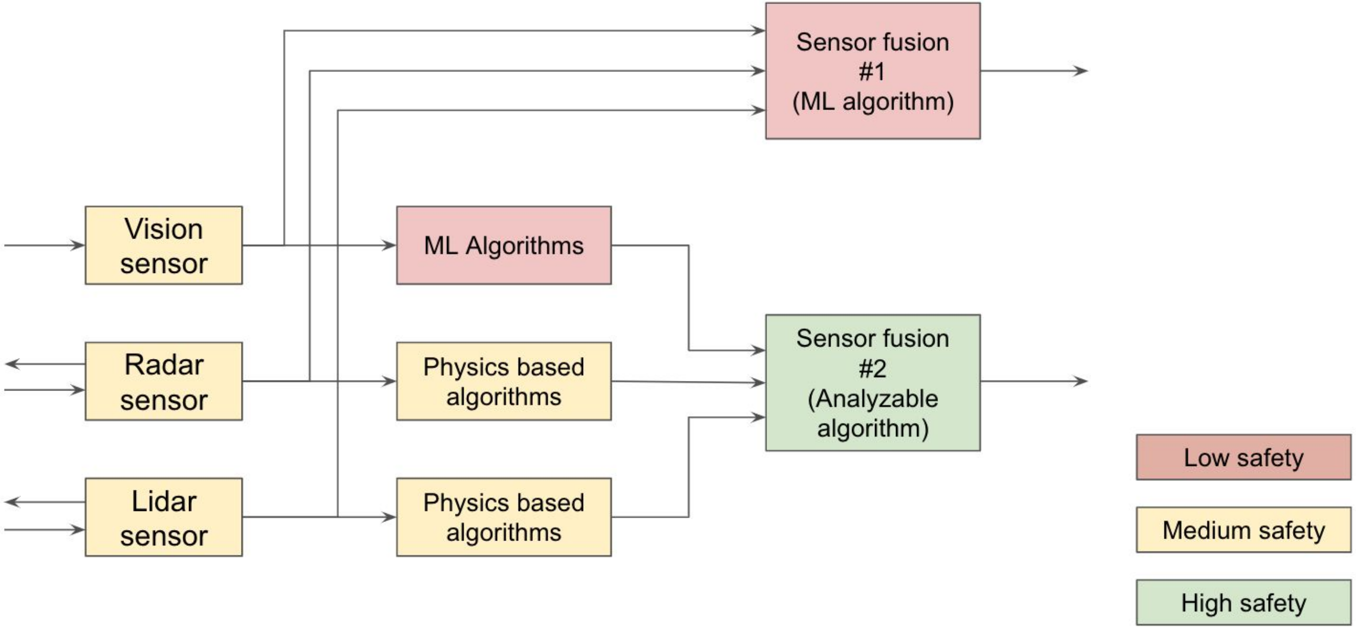
Example: Perception

From:

No false positive;
minimize false negative

To:

No false negative;
Minimize false positive



An example architecture

Prediction: AI-heavy vs Physics?

Semantic perception: Based on classification and behavior prediction in context.

Physics: Newtonian mechanics. Minimize energy of an impact and loss of driveable surface. Smaller time frames.

TOYOTA
RESEARCH INSTITUTE

Context: The Operational Design Domain

Context: Operational Design Domain (ODD)

- Roughly: Conditions for AV function to operate
- Safety description must be accompanied by ODD description

⇒ For L4 functions, the ODD must be "knowable" to the AV function

- Observable, inferable, accessible

"The ODD excludes heavy rain" ← Poor formulation if AV can not know what heavy rain is, or that it is happening.

Create an ODD in four simple(?) steps

1. Define all 'Concepts' that you care about
    a. Concepts have 'Properties' and Properties have 'Values'

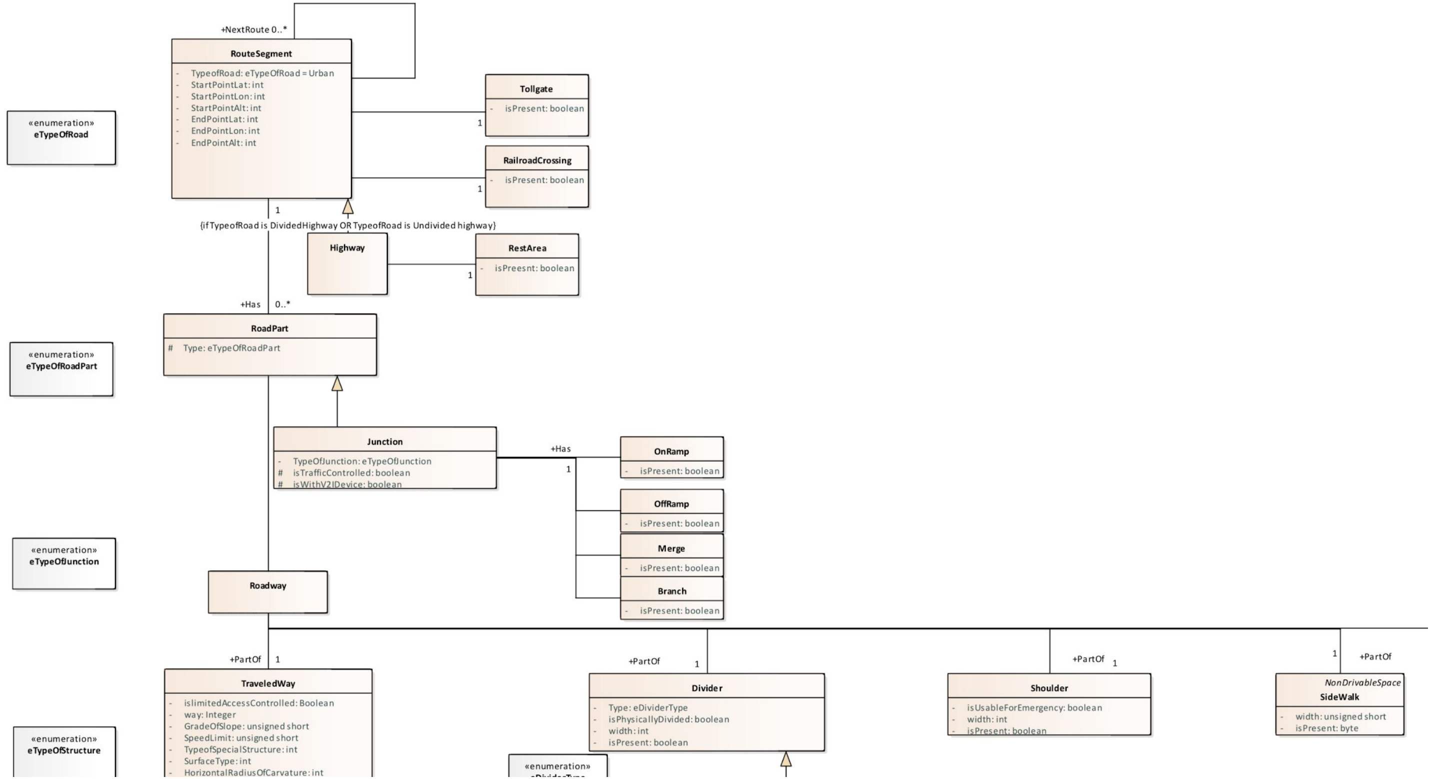
2. Organize the concepts into a 'Hierarchy' suitable for your function

3. Create 'Relationships of interest'
    a. Between Concepts
    b. Between Properties of Concepts

4. Define constraints on Concept PropertyValues and Relationships

Example ODD fragment

Physical Infrastructure(ODD Ontology view)



+NextRoute 0..*

RouteSegment
- TypeofRoad: eTypeOfRoad = Urban
- StartPointLat: int
- StartPointLon: int
- StartPointAlt: int
- EndPointLat: int
- EndPointLon: int
- EndPointAlt: int

«enumeration»
eTypeOfRoad

Tollgate
- isPresent: boolean

1

RailroadCrossing
- isPresent: boolean

1

1

{if TypeofRoad is DividedHighway OR TypeofRoad is Undivided highway}

Highway

RestArea
- isPreesnt: boolean

1

+Has 0..*

RoadPart
Type: eTypeOfRoadPart

«enumeration»
eTypeOfRoadPart

Junction
- TypeOfJunction: eTypeOfJunction
isTrafficControlled: boolean
isWithV2IDevice: boolean

+Has
1

OnRamp
- isPresent: boolean

OffRamp
- isPresent: boolean

Merge
- isPresent: boolean

Branch
- isPresent: boolean

«enumeration»
eTypeOfJunction

Roadway

+PartOf 1

TraveledWay
- islimitedAccessControlled: Boolean
- way: Integer
- GradeOfSlope: unsigned short
- SpeedLimit: unsigned short
- TypeofSpecialStructure: int
- SurfaceType: int
- HorizontalRadiusOfCarvature: int

«enumeration»
eTypeOfStructure

+PartOf 1

Divider
- Type: eDividerType
- isPhysicallyDivided: boolean
- width: int
- isPresent: boolean

«enumeration»

+PartOf 1

Shoulder
- isUsableForEmergency: boolean
- width: int
- isPresent: boolean

1 +PartOf

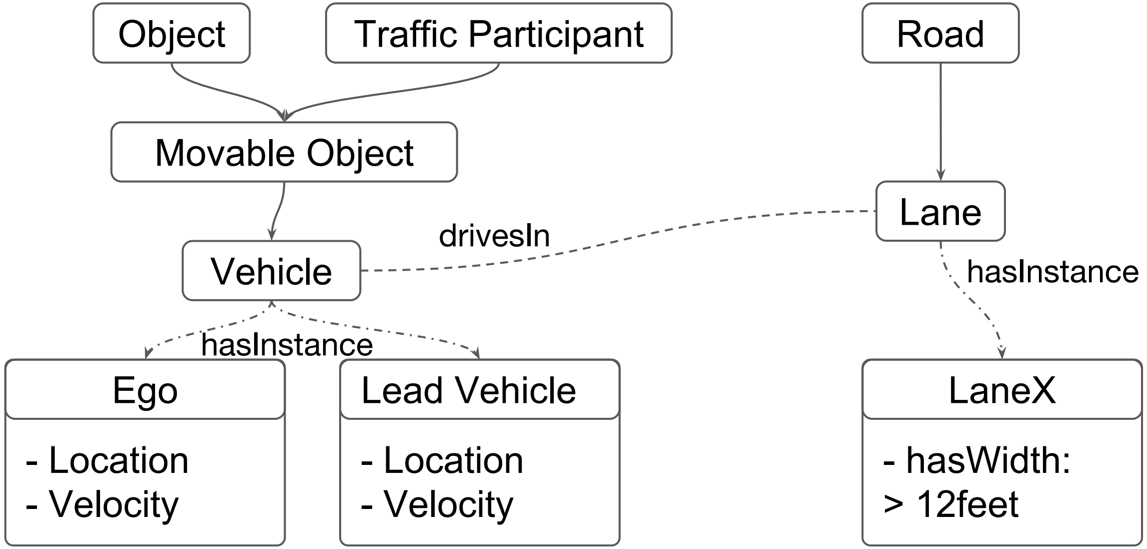
NonDrivableSpace
SideWalk
- width: unsigned short
- isPresent: byte

Ontologies: backbone of ODD and Safety

1 Concepts:: Properties: PropertyValues

2 Concept hierarchy

3 Relationships of interest



ODD: Relevant and knowable subset of Ontology

ODD Instance: Constraints on Concept PropertyValues

Safety: Constraints on Concept PropertyValues and Relationships

Always

$$\frac{\text{Ego.location - Lead.location}}{\text{Ego.velocity - Lead.velocity}} > 2s$$

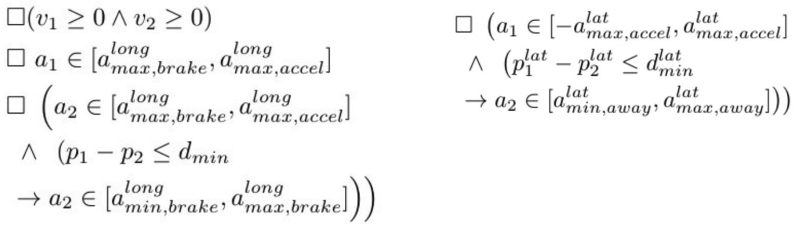
Synthesis of Ontology-based safety monitors

Safety: Constraints on Concept PropertyValues and Relationships in Ontology

Plain text:

- Don't drive backwards; keep acceleration and braking within bounds
- Maintain "safe distance" from lead vehicle
- Stay within a margin of lane boundaries

Formal rules:



$$\square(v_1 \geq 0 \wedge v_2 \geq 0)$$

$$\square \; a_1 \in [a^{long}_{max,brake}, a^{long}_{max,accel}]$$

$$\square \; \Big(a_2 \in [a^{long}_{max,brake}, a^{long}_{max,accel}]$$

$$\wedge \;\; (p_1 - p_2 \leq d_{min}$$

$$\rightarrow a_2 \in [a^{long}_{min,brake}, a^{long}_{max,brake}]\Big)\Big)$$

$$\square \;\; (a_1 \in [-a^{lat}_{max,accel}, a^{lat}_{max,accel}]$$

$$\wedge \;\; (p^{lat}_1 - p^{lat}_2 \leq d^{lat}_{min}$$

$$\rightarrow a_2 \in [a^{lat}_{min,away}, a^{lat}_{max,away}]))$$

Remarks:

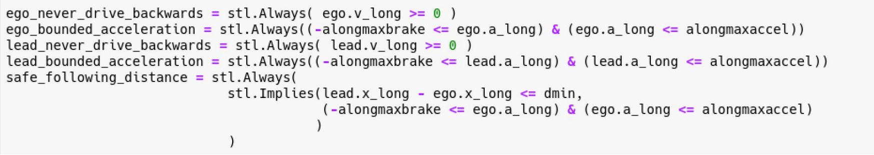
Requirements become first class software objects

Executable, Maintainable

Formal logic unlocks

Falsification, conformance of subsystems with systems

Code:

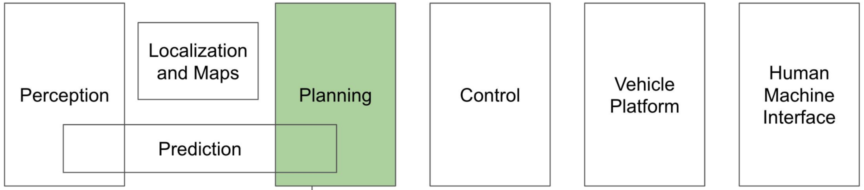


```
ego_never_drive_backwards = stl.Always( ego.v_long >= 0 )
ego_bounded_acceleration = stl.Always((-alongmaxbrake <= ego.a_long) & (ego.a_long <= alongmaxaccel))
lead_never_drive_backwards = stl.Always( lead.v_long >= 0 )
lead_bounded_acceleration = stl.Always((-alongmaxbrake <= lead.a_long) & (lead.a_long <= alongmaxaccel))
safe_following_distance = stl.Always(
                            stl.Implies(lead.x_long - ego.x_long <= dmin,
                                        (-alongmaxbrake <= ego.a_long) & (ego.a_long <= alongmaxaccel)
                                       )
                           )
```

TOYOTA
RESEARCH INSTITUTE

Coverage and Residual Risk

Coverage and residual risk



Perception

Localization and Maps

Planning

Control

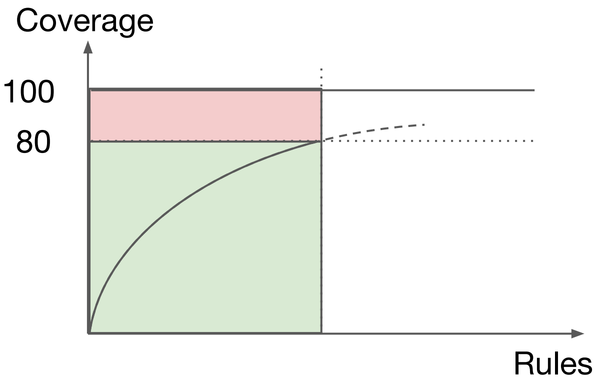
Vehicle Platform

Human Machine Interface

Prediction

- Find a finite set of planning rules
- Adherence to rules shall avoid fatal incidents
- Prove that AV system will not violate rules

For a defined ODD



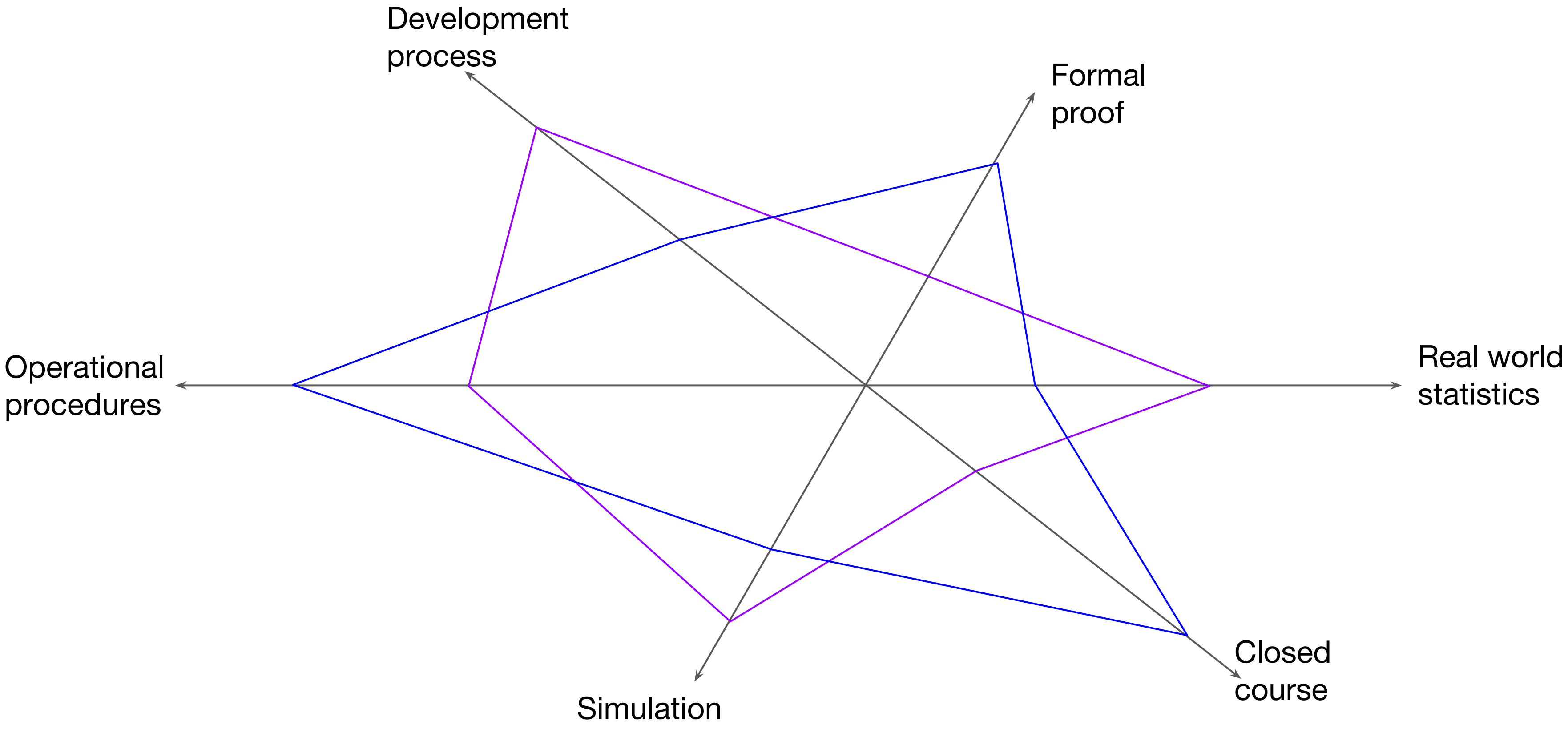
In a given ODD:

Coverage: What percentage of undesired outcomes would be avoided by selected set of safety rules?
Residual risk: For a given system implementation, what is probability of safety rule violation?

TOYOTA
RESEARCH INSTITUTE

Methods of evidence

Methods of evidence



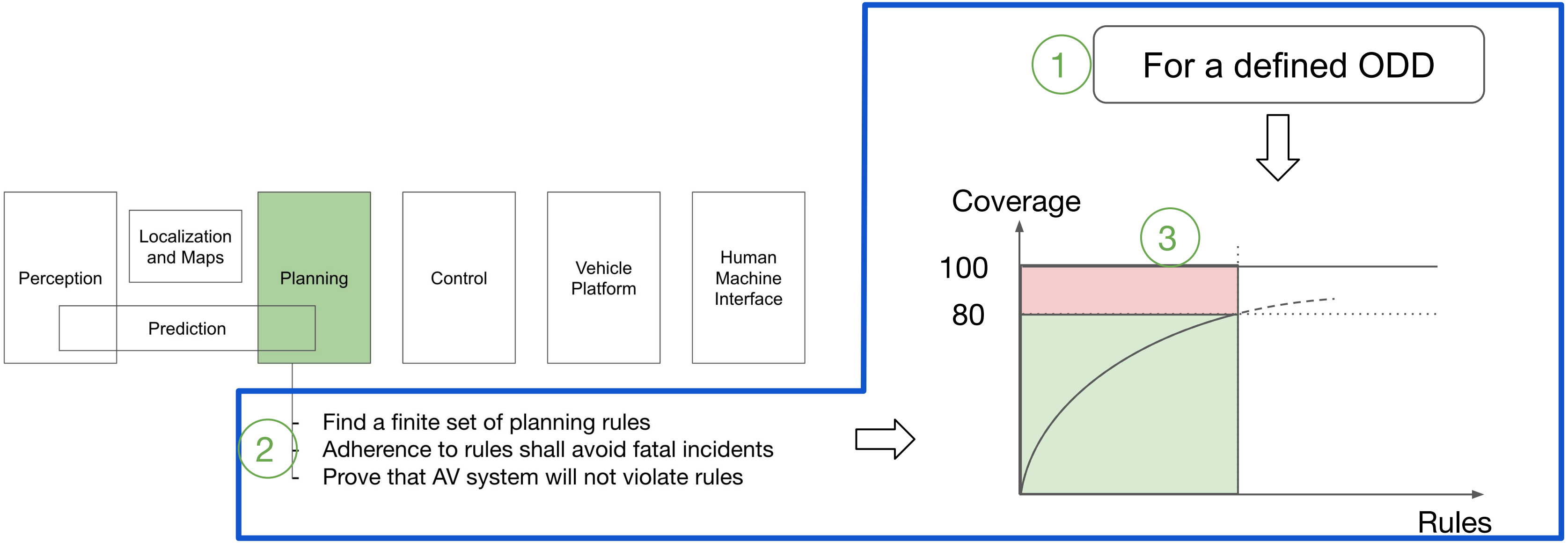
The bigger questions

How safe is safe enough?

- What are the metrics?
- Who decides?
- If acceptable values are found for each safety metric, how do you know your system is achieving those metrics?
- Comparisons with human drivers?

Alternatively: Can you calculate the probability of violation of safety rules for a given system implementation?

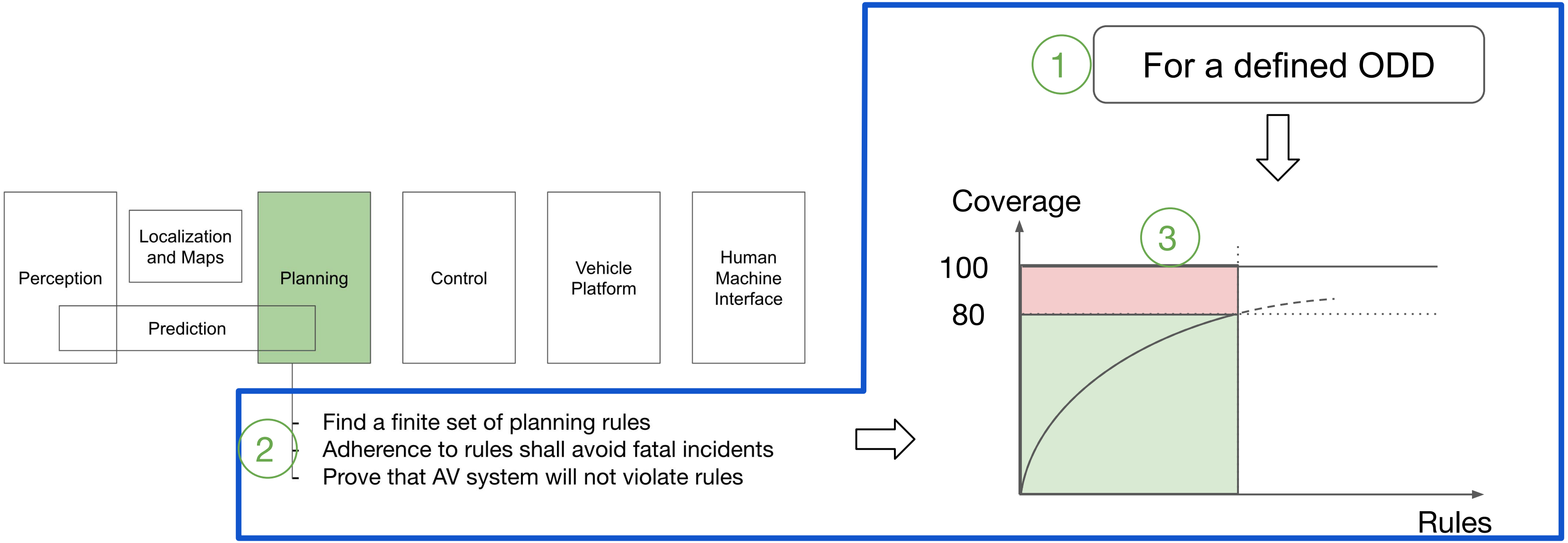
Three areas for cooperation



- Find a finite set of planning rules
- Adherence to rules shall avoid fatal incidents
- Prove that AV system will not violate rules

Bonus area: Assumptions within ODD?

Data sharing



1. (Abstracted) Data showing that set of safety rules need adjustments/additions
2. (Abstracted) Data showing that the coverage in an ODD needs to be adjusted

Content of an AV Safety case

1	PHILOSOPHY
2	CONTEXT
3	DESIGN, IMPLEMENTATION
4	EVIDENCE
5	COVERAGE/RESIDUAL RISK
6	LARGER QUESTIONS

A credible AV safety case must provide rational evidence-based argumentation for each area

Thank You