ICT技術の信頼を支える
社会インフラとしてのソフトウェア科学

国立情報学研究所
総合研究大学院大学
ERATO 蓮尾メタ数理システムデザインプロジェクト
蓮尾 一郎

情報学科・専攻協議会 2023/07/15

Outline

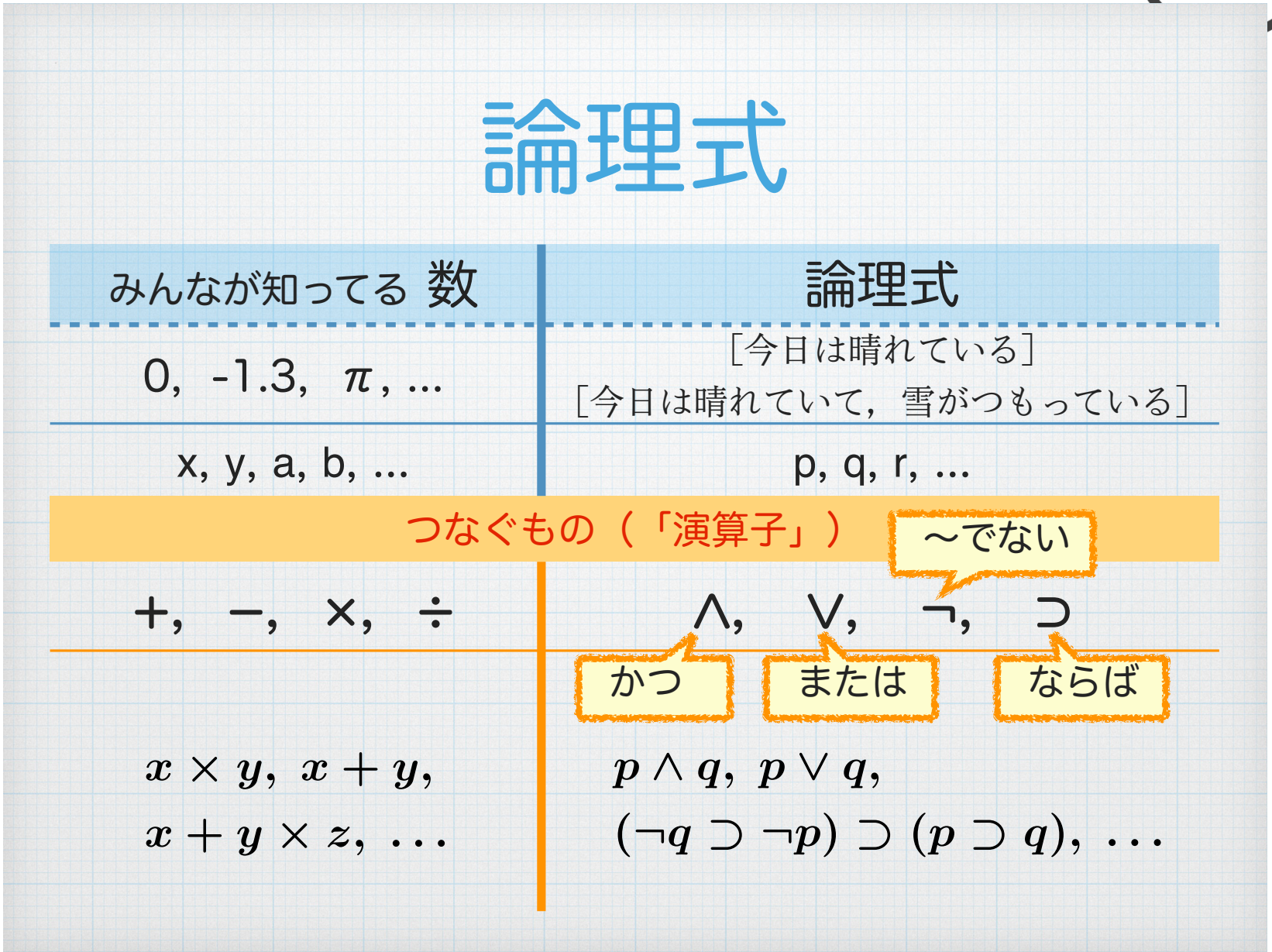
- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ

- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]

- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ

論理学とは

- ソフトウェア科学は**論理学**に立脚
- 高校生向けアウトリーチスライドから抜粋
  （言葉遣いご容赦ください）



論理式を計算する

みんなが知ってる 数	論理式

計算のルール

$$x + y = y + x$$

$$x \times y = y \times x$$

$$x \times (y + z) = x \times y + x \times z$$

$$(x + y)^2 = x^2 + 2xy + y^2$$

p が本当で

p ⊃ q が本当な

$$p \wedge q$$

$$\frac{p}{p \vee q}$$

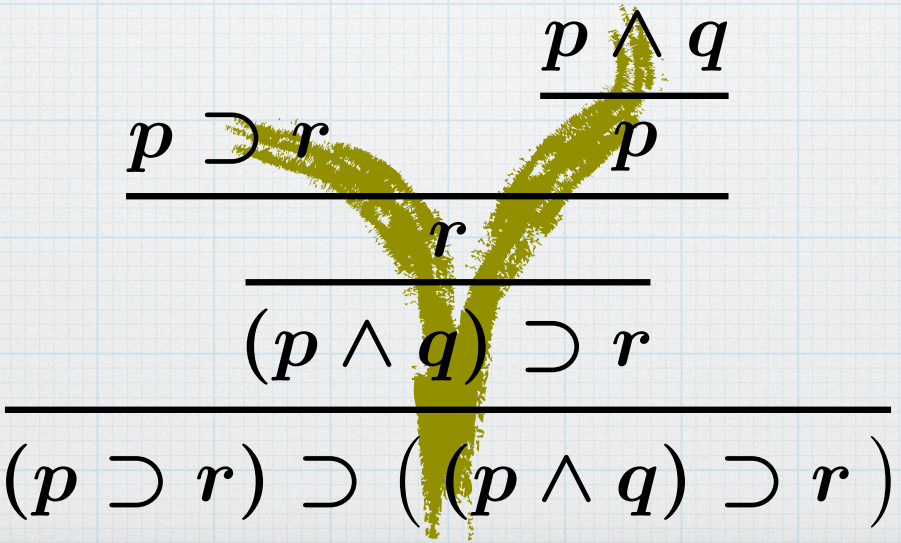
$$\frac{p \qquad p \supset q}{q}$$

p

q が本当

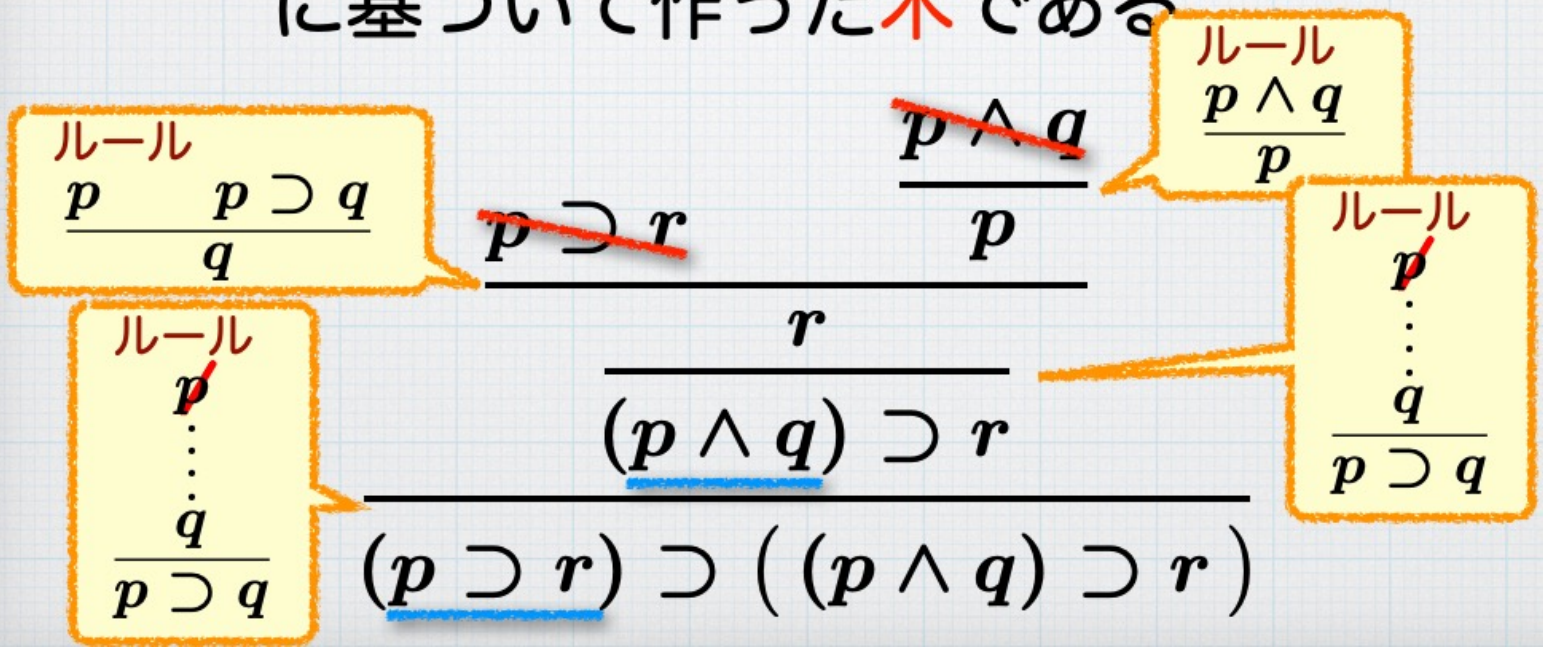
証明の数学的な定義

* 定義
  証明とは，さっきの計算のルールに基づいて作った木である．



証明の数学的な定義

* 定義
  証明とは，さっきの計算のルールに基づいて作った木である

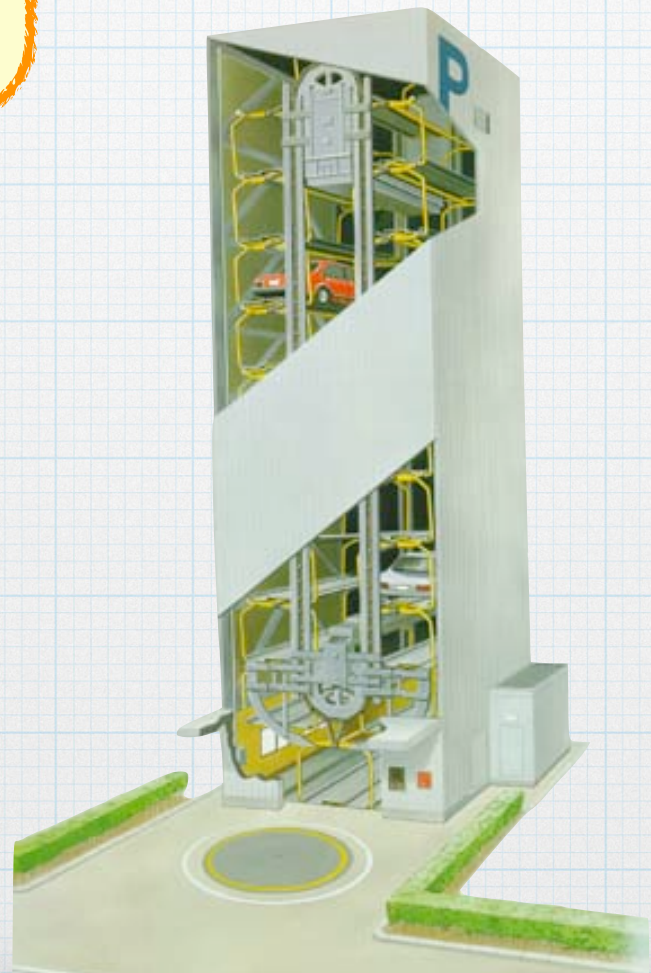


ビジネスでの証明

オレが責任持つっす！

絶対大丈夫っす！

この駐車場，
どうっすか！？



・・・
（ダメだこいつ）

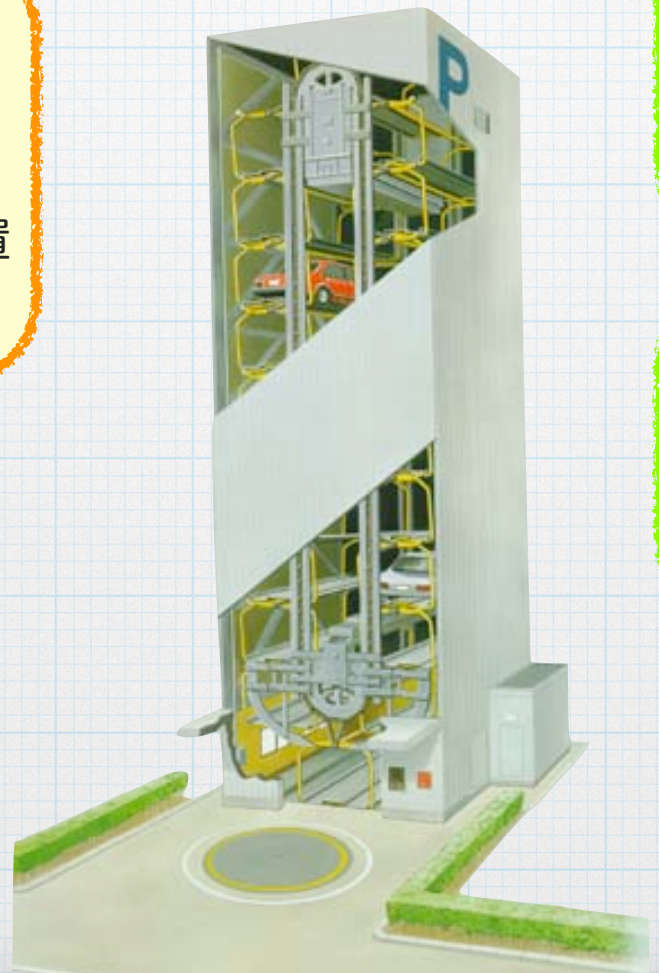
ホントに？ なんで？

ゴンドラが衝突したりしない？ 大丈夫？

ビジネスでの証明

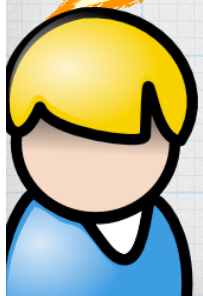
はい，大丈夫です．
なぜなら，任意の状態 **s** に対して，ゴンドラ g_1 の位置を x_1 とすると，．．．

この駐車場，
どうでしょう．

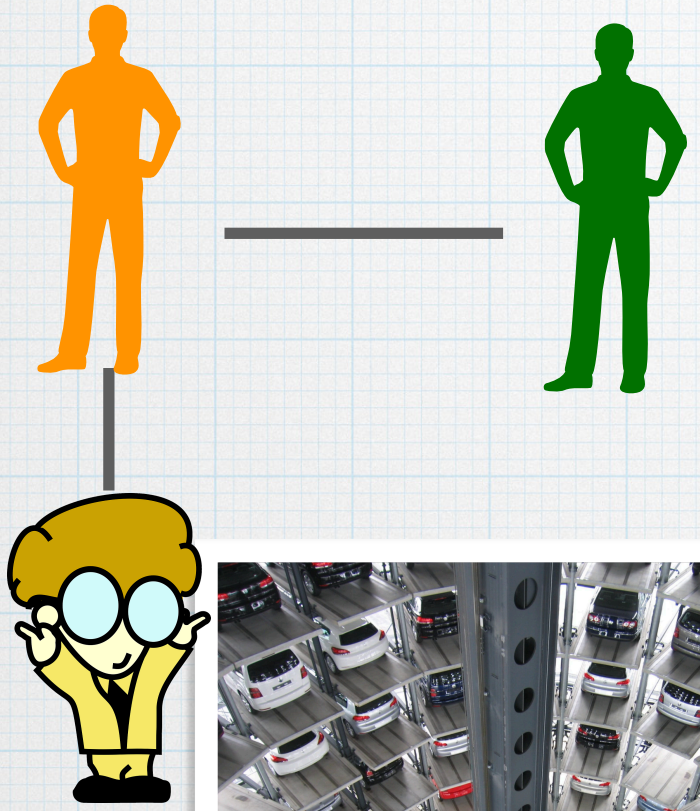


．．．（読んでる）
なるほど．それでは1基いただこう．

ゴンドラが衝突したりしない？ 大丈夫？



証明で金もうけ



https://www.needpix.com/photo/178834/nerd-cartoon-geek-character-glasses-isolated-funny-face-smart

* 定理証明のコスト：
  * たとえば，ヨーロッパの計算機科学の博士課程（有給，3–4年）
  * 1000万円をゆうに超える

高校生用の啓蒙スライド「証明問題がんばろう」

Outline

- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ

- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]

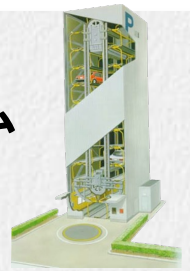
- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ

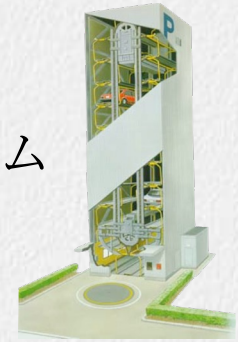
証明のためには定義が必要

定義 **完備距離空間**とは任意のコーシー列が極限を持つ距離空間のことを言う．

定理 完備距離空間 X 上の収縮写像 $f: X \to X$ は不動点を持つ．さらに，この不動点は一意に定まる．

証明． $x \in X$ を任意に選び，点列 $x, f(x), f(f(x)), \ldots$ を考えると，f が収縮写像であることよりこれはコーシー列．よって完備性の定義より極限 x_0 を持つ．$x_0 = f(x_0)$ であることは容易に示される．□

定義 **駐車場システム**  とは…??

定理 駐車場システム  は安全である．

証明． ???

- 数学的議論の対象（登場人物）はすべて厳密に定義されなければならない
- 定義は数学的に正確で，かつ本質をとらえた単純なものでなければならない（単純でないと使いにくい）
- ➔ 数理モデリング！

The Modeling Problem in Emerging ICT

- Theorems need *definitions*;
  formal verification needs *modeling*

Emerging ICT

Conventional ICT Systems

- They operate following a logical recipe (= program)
- Programs are mathematical models

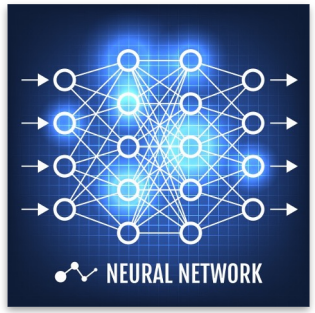
✓

Cyber-Physical Systems

- E.g. automated driving cars
- Physical components? Other cars? Pedestrians?

??

Statistical AI Systems



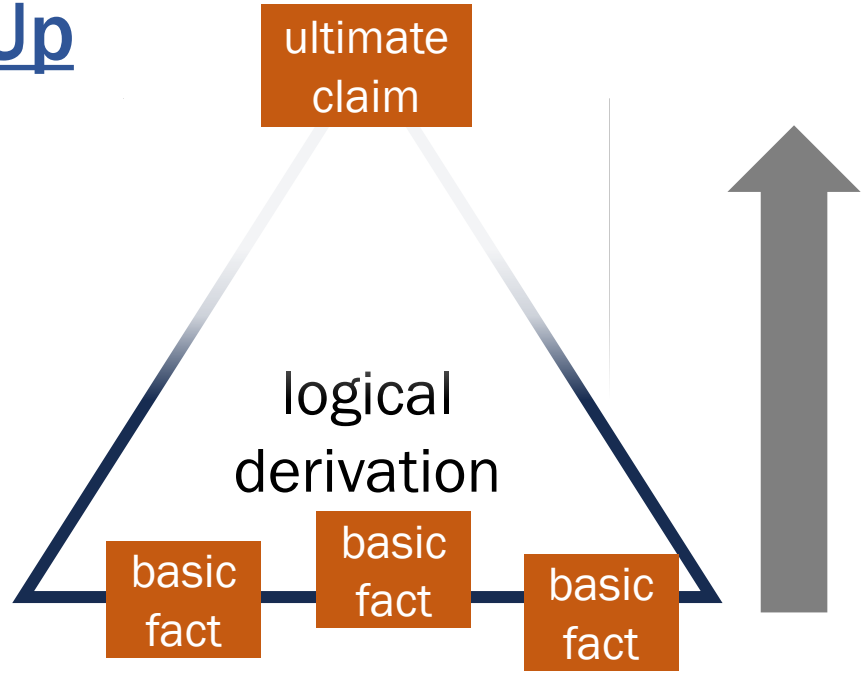
- They decide using matrices learned from big data
- Those matrices are too big and fragile for logical analysis

??

...

Use of Logic—Bottom Up or Top Down?

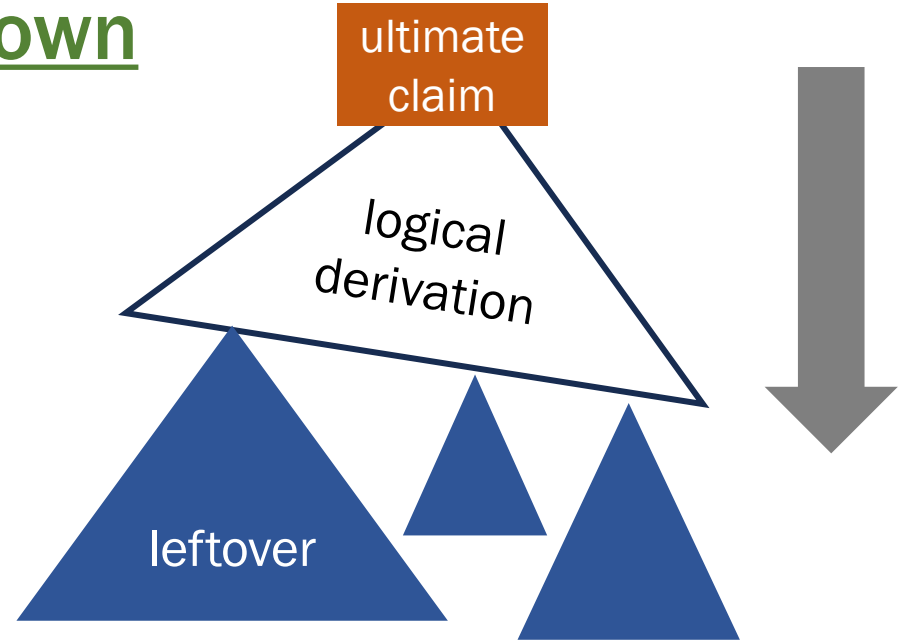
Bottom Up



Conventional use of logic

- ✓ Start from basic & unquestionable facts, and build up unquestionable facts
- ✗ Basic facts are often not available (e.g. models are rare for cyber-physical systems)
- ✗ The ultimate claim is far, far away
- ✗ Zero value in incomplete proofs

Top Down



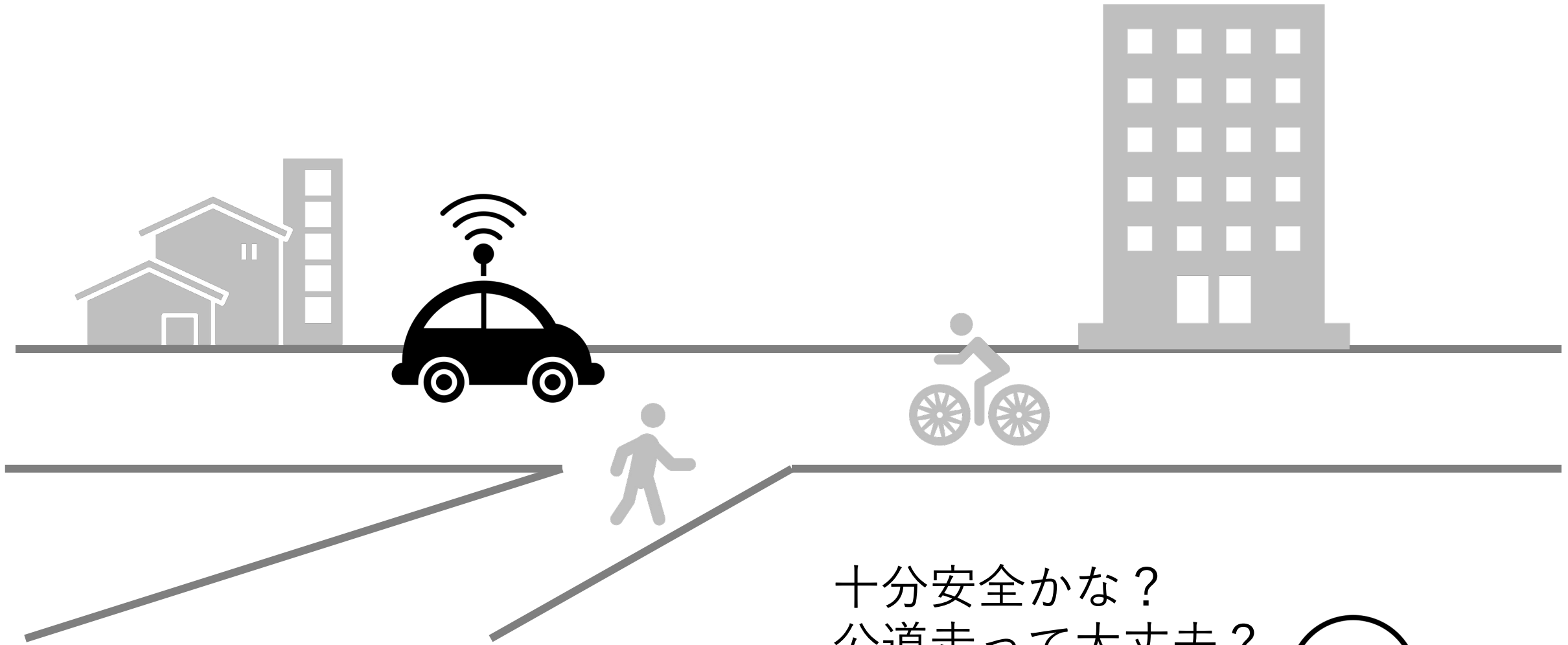
- ✗ The ultimate claim's validity is only conditional
- ✓ Scalable, best-effort (Dig as deep as the budget runs)
- ✓ Connects smoothly with statistical methods
- ✓ Handles black- & gray-box models (Test or runtime-monitor leftover assumptions)
- ✓ Logical explainability and traceability. "Logical safeguard"

Outline

- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ

- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]

- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ

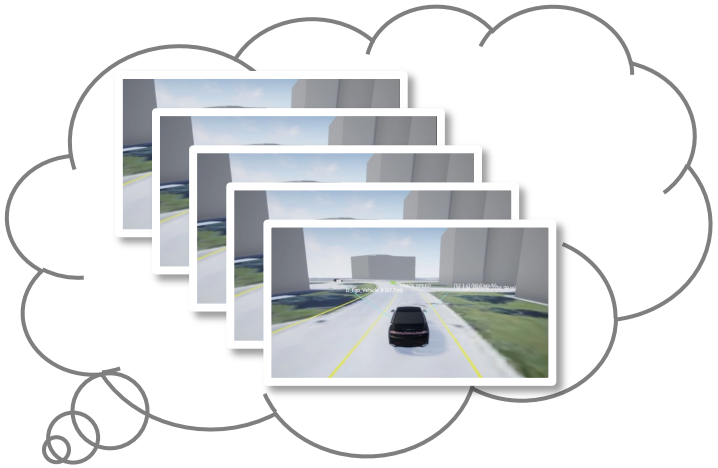


十分安全かな？
公道走って大丈夫？

事故統計による保証

9.1 crashes
per million miles
driven

テストとシミュレーションによる保証

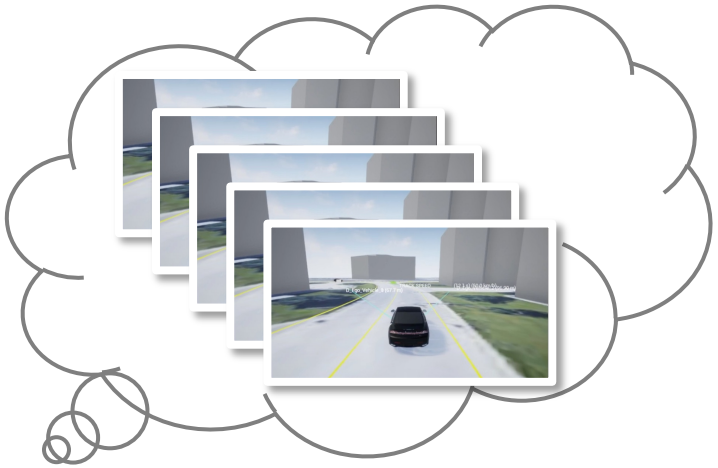


なぜこれで十分と言える？

事故統計による保証

9.1 crashes
per million miles
driven

テストとシミュレーションによる保証



説明責任を果たせる？

Proof.

We prove the first statement. The rest is shown symm

Let $S \subseteq L$ be an arbitrary subset. We let $S^{\downarrow}$ be th
that is,

$$S^{\downarrow} \coloneqq \{y \in L \mid y \sqsubseteq s \text{ for each } s \in$$

Since $S^{\downarrow} \subseteq L$ is a subset of L, it has its supremum $\bigsqcup S^{\downarrow}$ in the semilattice $(L, \sqsubseteq)$. We claim that $\bigsqcup S^{\downarrow}$ is the infimum

To prove the claim, it suffices to show the two-way acterization in (2.1), that is, we need to show

$$\frac{y \sqsubseteq s \quad \text{for each } s \in S}{y \sqsubseteq \bigsqcup S^{\downarrow}}.$$

For the downward implication in **??**,

$$\begin{aligned} & y \sqsubseteq s \quad \text{for each } s \in S \\ \Longrightarrow \quad & y \in S^{\downarrow} && \text{by def. of } S^{\downarrow} \\ \Longrightarrow \quad & y \sqsubseteq \bigsqcup S^{\downarrow} && \text{since } \bigsqcup S^{\downarrow} \text{ is an u} \end{aligned}$$

For the upward implication in **??**, we first observe

$$\bigsqcup S^{\downarrow} \sqsubseteq s \quad \text{for each } s \in S.$$


12 : 00 PM
SPEED - 0 km/h
ENGINE STATUS - OFF
New Users
Season
Comparison of sessions
Views
Unlock
Child lock
7.1 gal
1.3 gal
0.9 gal
84 psi
95%

Responsibility-Sensitive Safety (責任感知型安全論，RSS)

[Shalev-Shwartz et al., arXiv preprint, 2017]

補題（条件付き安全性補題）
各車が RSS ルールを守れば，
衝突は起きない

数学的に証明

＋

仮定（ルール遵守仮定）
各車は RSS ルールを守る．

各車（のメーカー）が責任を持つ

→

定理（安全性定理）
衝突は起きない

Responsibility-Sensitive Safety (責任感知型安全論，RSS)

[Shalev-Shwartz et al., arXiv preprint, 2017]

非常に賢い「割り切り方」

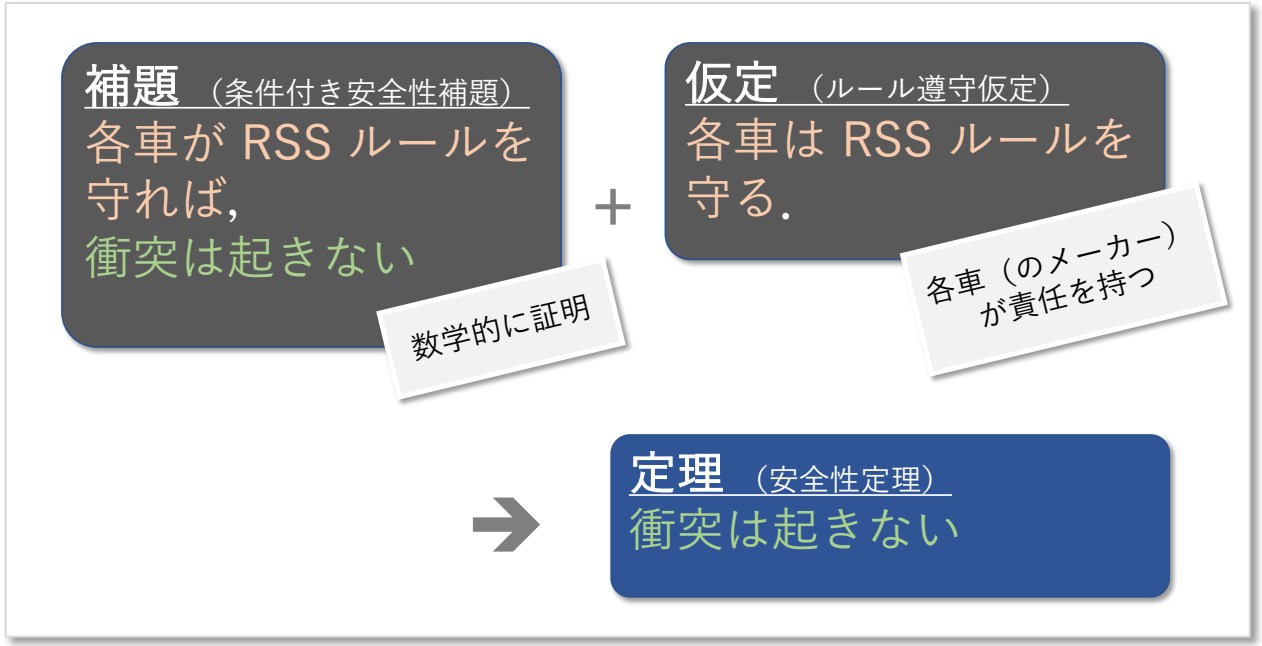
<u>条件付き安全性補題：</u>

- 複雑すぎず，
  数学的証明が現実的に可能

<u>ルール遵守仮定：</u>

一見巨大な仮定のようだが，

- 社会的契約（法規制，国際規格，業界標準）
  として要請するに適した粒度
- 監視＆介入による強制が可能
  （安全アーキテクチャ，後述）



RSS ルールの例

一車線同方向運転シナリオにおける追突回避のためのルール

[Shalev-Shwartz et al., arXiv preprint, 2017]

RSS 条件：
車間距離を

$$d_{\min} = \left[v_r\,\rho + \frac{1}{2} a_{\max,\mathrm{accel}}\,\rho^2 + \frac{(v_r + \rho\, a_{\max,\mathrm{accel}})^2}{2a_{\min,\mathrm{brake}}} - \frac{v_f^2}{2a_{\max,\mathrm{brake}}} \right]_+$$

以上確保すること

適切反応 (proper response)：
上記RSS 条件の違反が予見される場合,
反応時間 ρ 以内に 減速度 $a_{\min,\mathrm{brake}}$ でブレーキすること

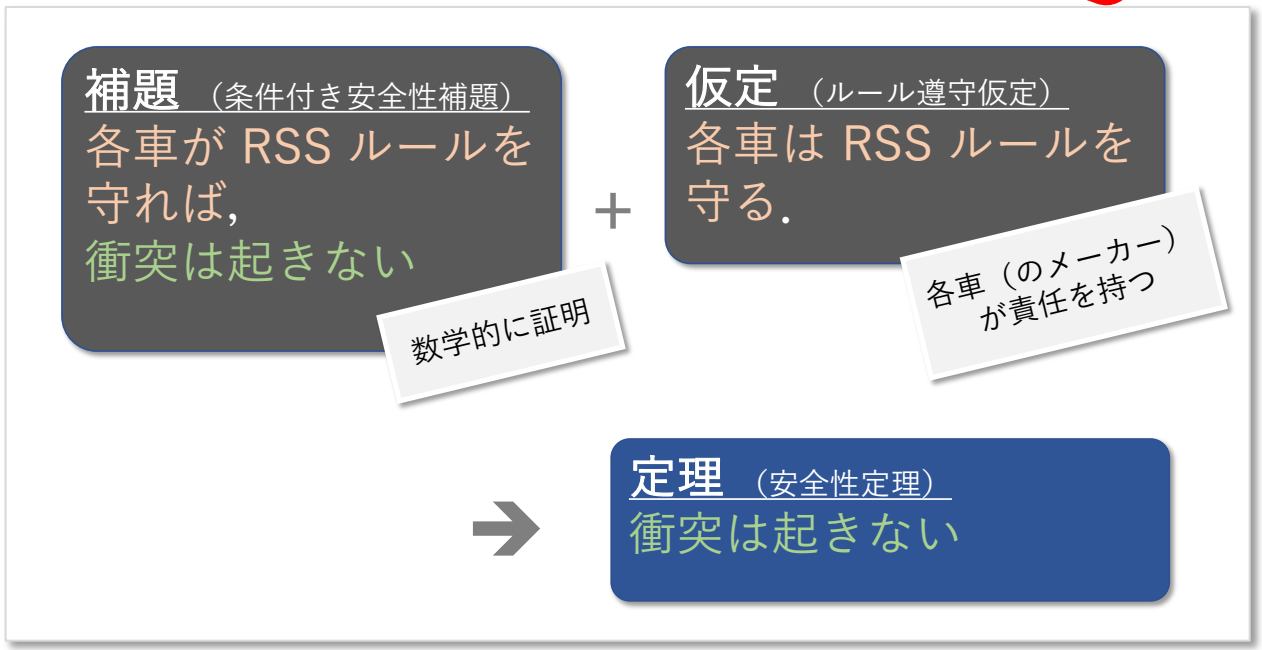
条件付き安全性補題：
RSS 条件が真である状態から適切反応を実行すれば，衝突は発生しない

Responsibility-Sensitive Safety (責任感知型安全論，RSS)

再掲

[Shalev-Shwartz et al., arXiv preprint, 2017]

非常に賢い「割り切り方」



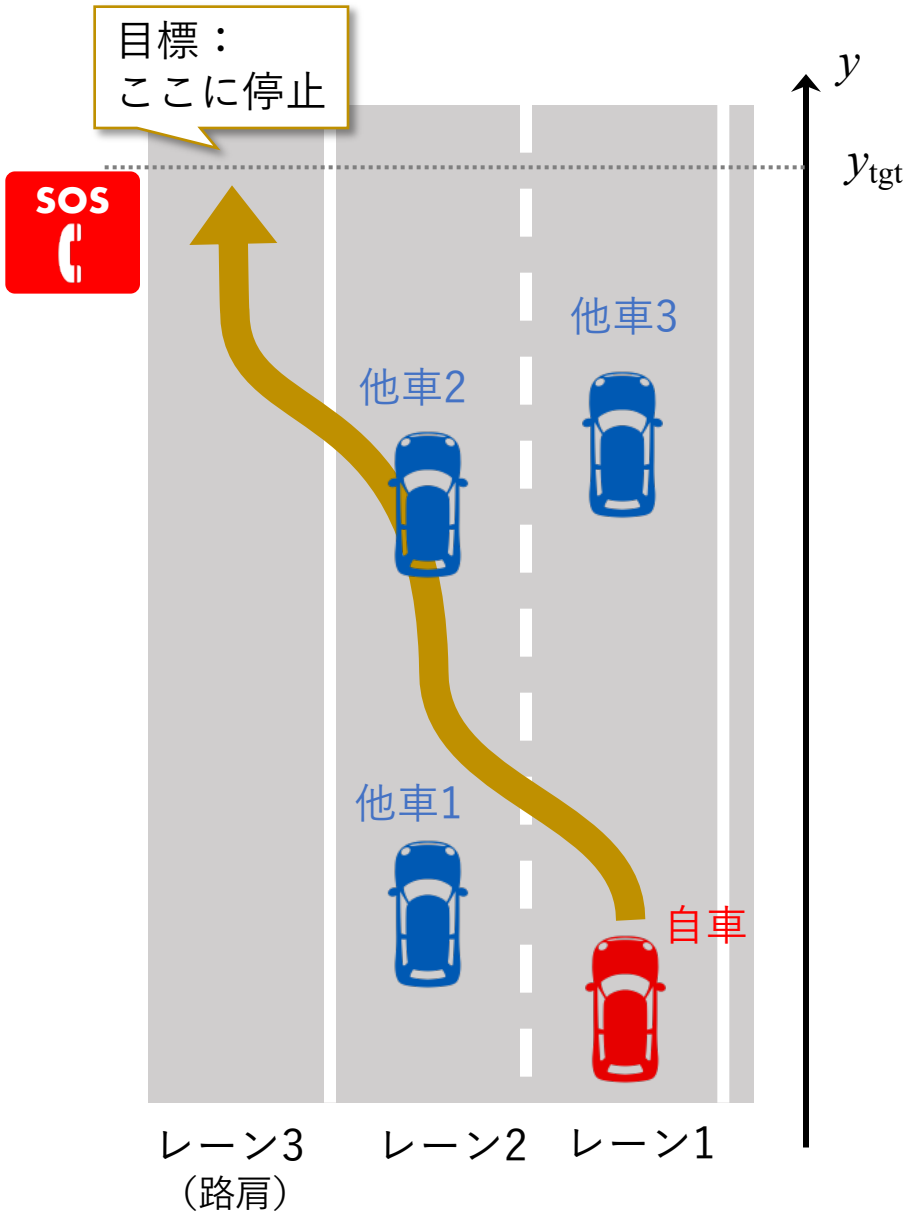
条件付き安全性補題：

- 複雑すぎず，
  数学的証明が現実的に可能

ルール遵守仮定：

一見巨大な仮定のようだが，

- 社会的契約（法規制，国際規格，業界標準）として要請するに適した粒度
- 監視＆介入による強制が可能
  （安全アーキテクチャ，後述）



- このような運転シナリオ
  （路肩停止）ではどうか？
  （ODD限定の自動運転車で，
  制御を人間に渡すために必須…）
- 他車１の前で合流？ 後で？
  追い越すために加速したら
  停止位置をオーバーランするかも？
  …

  → 複雑さが段違い

我々の成果：
RSS 証明の「形式化」による RSS ルールの本格展開

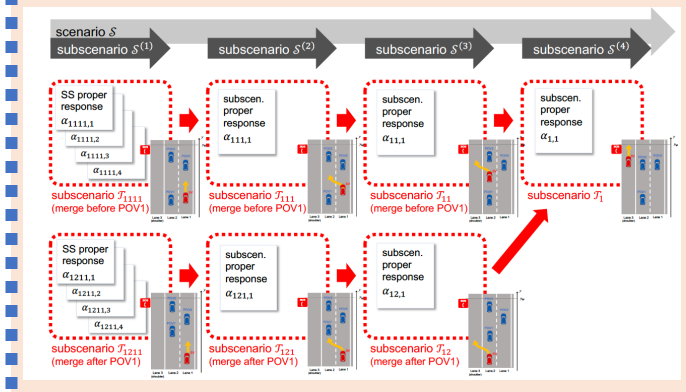
RSS
Responsibility-Sensitive Safety
（責任感知型安全論）
[Shalev-Shwartz et al., arXiv, 2017]
- 安全ルールの基本的方法論
- 国際規格化の動き
(IEEE 2846)
- 複雑なシナリオに対するルール策定・証明手法は未整備
- **特に，衝突回避以外の目的への対応事例がない**

微分プログラム論理 dFHL
（今回の成果）

$$\frac{\begin{array}{ll}\text{inv}: & A \Rightarrow e_{\text{inv}} \sim 0 \quad e_{\text{var}} \geq 0 \wedge e_{\text{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\text{inv}} \simeq 0 \\ \text{var}: & A \Rightarrow e_{\text{var}} \geq 0 \quad e_{\text{var}} \geq 0 \wedge e_{\text{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\text{var}} \leq e_{\text{ter}} \\ \text{ter}: & A \Rightarrow e_{\text{ter}} < 0 \quad e_{\text{var}} \geq 0 \wedge e_{\text{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\text{ter}} \leq 0\end{array}}{\{A\}\ \text{dwhile}\,(e_{\text{var}} > 0)\,\dot{\mathbf{x}} = \mathbf{f}\ \ \{e_{\text{var}} = 0 \wedge e_{\text{inv}} \sim 0\} : e_{\text{inv}} \sim 0 \wedge e_{\text{var}} \geq 0}\ (\text{DW}_{\text{H}})$$

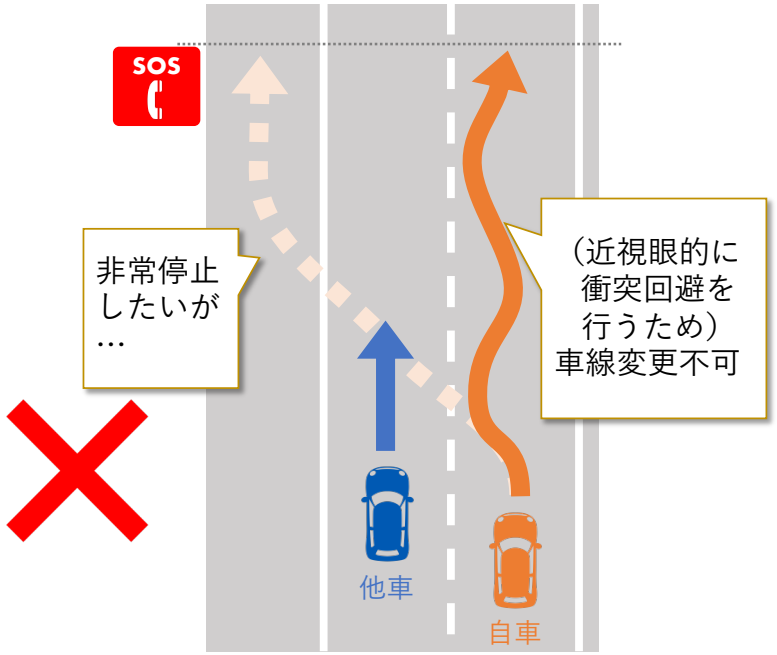
- 安全ルール導出・証明のための論理体系

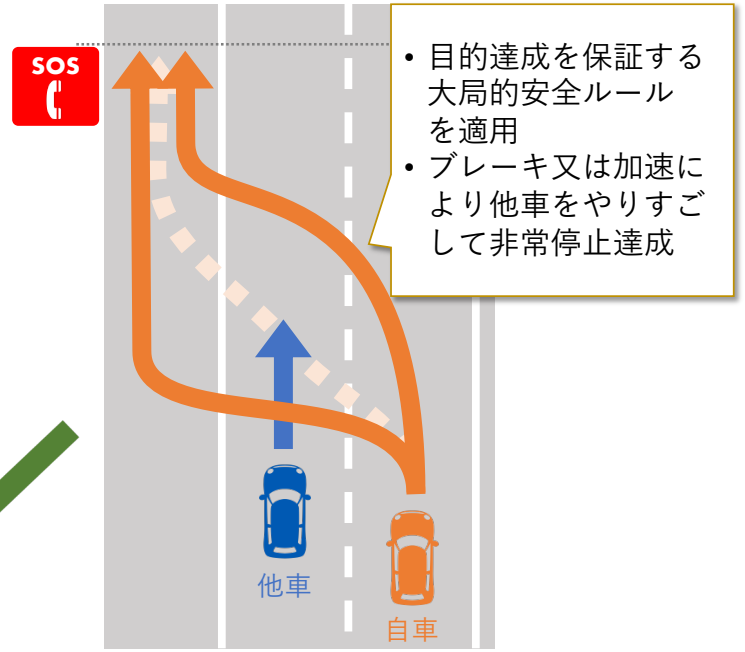
dFHL による逐次的推論・ルール導出ワークフロー
（今回の成果）



- 複雑な行動計画を分割，それぞれ論理的解析し，結果を結合
- 自動推論によるツールサポート

GA-RSS（今回の成果）
Goal-Aware
Responsibility-Sensitive Safety
[Hasuo+, IEEE T-IV, 2023]
- 衝突回避に加え，緊急停止等の**目的達成**もサポート
- 複数の行動を組み合わせた大局的安全ルール
- 現実の複雑な交通シナリオへの適用において必須



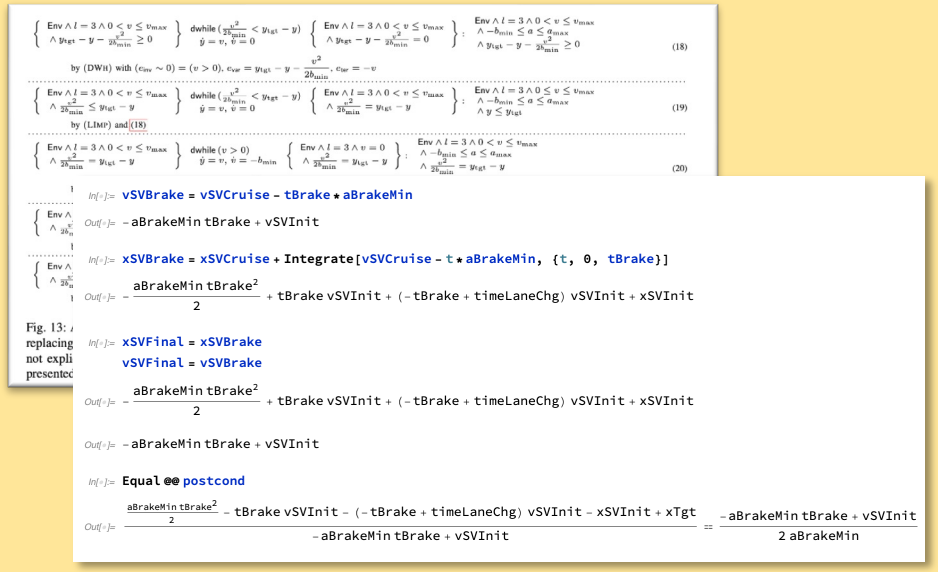


証明の形式化とは？

非形式的証明
（紙とペンで書く）

- 大量（数千ページとか），間違えやすい
- 後から or 他人による チェック・解析が困難

形式的証明
（コンピュータ上で書く）



```
In[•]:= vSVBrake = vSVCruise - tBrake * aBrakeMin
Out[•]= -aBrakeMin tBrake + vSVInit

In[•]:= xSVBrake = xSVCruise + Integrate[vSVCruise - t * aBrakeMin, {t, 0, tBrake}]
Out[•]= -(aBrakeMin tBrake^2)/2 + tBrake vSVInit + (-tBrake + timeLaneChg) vSVInit + xSVInit

In[•]:= xSVFinal = xSVBrake
        vSVFinal = vSVBrake
Out[•]= -(aBrakeMin tBrake^2)/2 + tBrake vSVInit + (-tBrake + timeLaneChg) vSVInit + xSVInit

Out[•]= -aBrakeMin tBrake + vSVInit

In[•]:= Equal @@ postcond
Out[•]= (aBrakeMin tBrake^2/2 - tBrake vSVInit - (-tBrake + timeLaneChg) vSVInit - xSVInit + xTgt)/(-aBrakeMin tBrake + vSVInit) == (-aBrakeMin tBrake + vSVInit)/(2 aBrakeMin)
```

- あらかじめ決まった証明言語と 書き換え規則で 証明の各ステップを記述（「記号推論」）
- 証明の正しさのチェックが ソフトウェア（proof checker）により可能

Differential program logic dFHL

- Hoare logic
  + ODEs (dwhile)
  + "safety condition"

$$\{A\}\ \alpha\ \{B\} : S$$

postcondition ↑ (true at the end of α)

"safety condition" ↑ (true throughout α)

- Reasoning about ODEs via differential invariants (barrier cert.) and ranking/Lyapunov functions
- Theoretically not so much different from Platzer's dL.
  Simplified, aiding proof engineers

Def. (dFHL programs)

$$\alpha, \beta \quad ::= \quad \mathsf{skip} \mid \alpha; \beta \mid x := e \mid \mathsf{if}\,(A)\,\alpha\,\mathsf{else}\,\beta \mid \mathsf{while}\,(A)\,\alpha \mid \mathsf{dwhile}\,(A)\,\{\,\dot{\mathbf{x}} = \mathbf{f}\,\}\,.$$

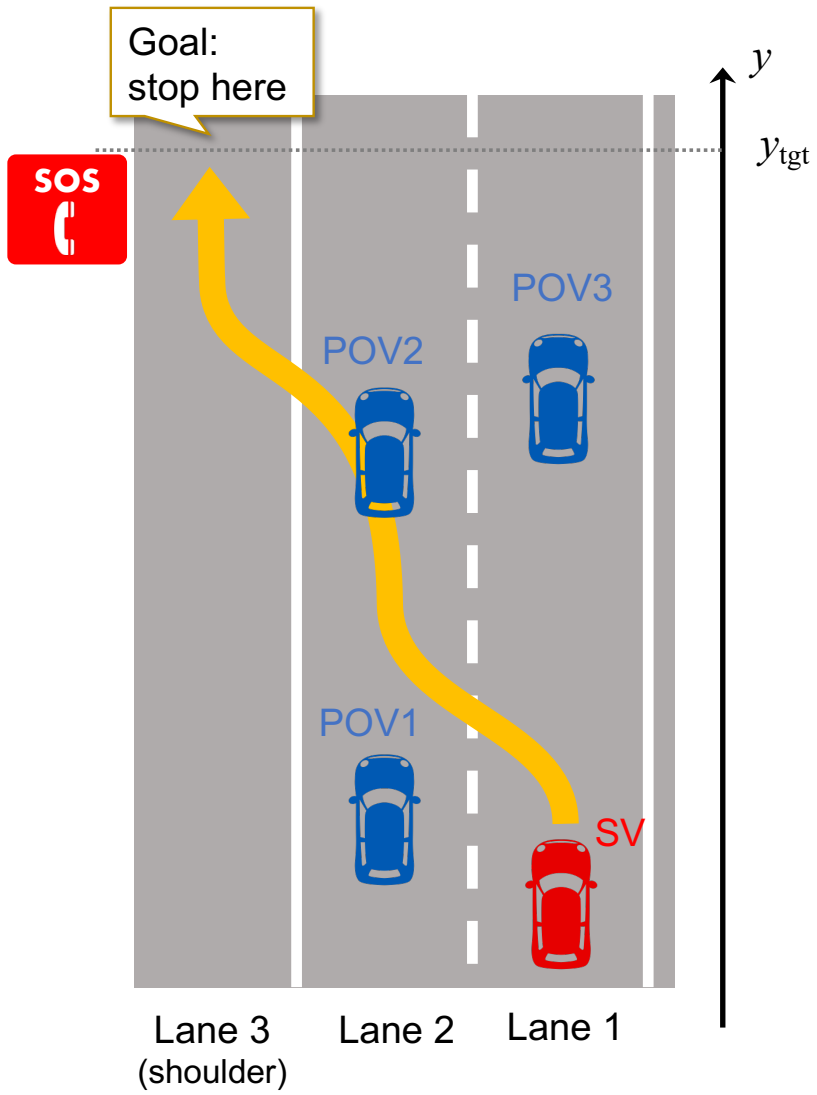
Def. (dFHL rules)

$$\frac{\{A\}\ \alpha\ \{B\} : S \qquad \{B\}\ \beta\ \{C\} : S}{\{A\}\ \alpha;\beta\ \{C\} : S}\ (\textsc{Seq})$$

$$\frac{\{A'\}\ \alpha\ \{B'\} : S' \qquad \begin{array}{c} A \Rightarrow A' \\ S' \wedge B' \Rightarrow B \\ S' \Rightarrow S \end{array}}{\{A\}\ \alpha\ \{B\} : S}\ (\textsc{Limp})$$

$$\frac{\begin{array}{ll} \mathsf{inv}: & A \Rightarrow e_{\mathsf{inv}} \sim 0 \quad e_{\mathsf{var}} \geq 0 \wedge e_{\mathsf{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\mathsf{inv}} \simeq 0 \\ \mathsf{var}: & A \Rightarrow e_{\mathsf{var}} \geq 0 \quad e_{\mathsf{var}} \geq 0 \wedge e_{\mathsf{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\mathsf{var}} \leq e_{\mathsf{ter}} \\ \mathsf{ter}: & A \Rightarrow e_{\mathsf{ter}} < 0 \quad e_{\mathsf{var}} \geq 0 \wedge e_{\mathsf{inv}} \sim 0 \Rightarrow \mathcal{L}_{\dot{\mathbf{x}}=\mathbf{f}}\, e_{\mathsf{ter}} \leq 0 \end{array}}{\{A\}\ \mathsf{dwhile}\,(e_{\mathsf{var}} > 0)\,\dot{\mathbf{x}} = \mathbf{f}\ \{e_{\mathsf{var}} = 0 \wedge e_{\mathsf{inv}} \sim 0\} : e_{\mathsf{inv}} \sim 0 \wedge e_{\mathsf{var}} \geq 0}\ (\textsc{DWh})^{\dagger}$$

Compositional Rule Derivation



- We shall derive

$$\{A\}\ \alpha\ \{B\} : S$$

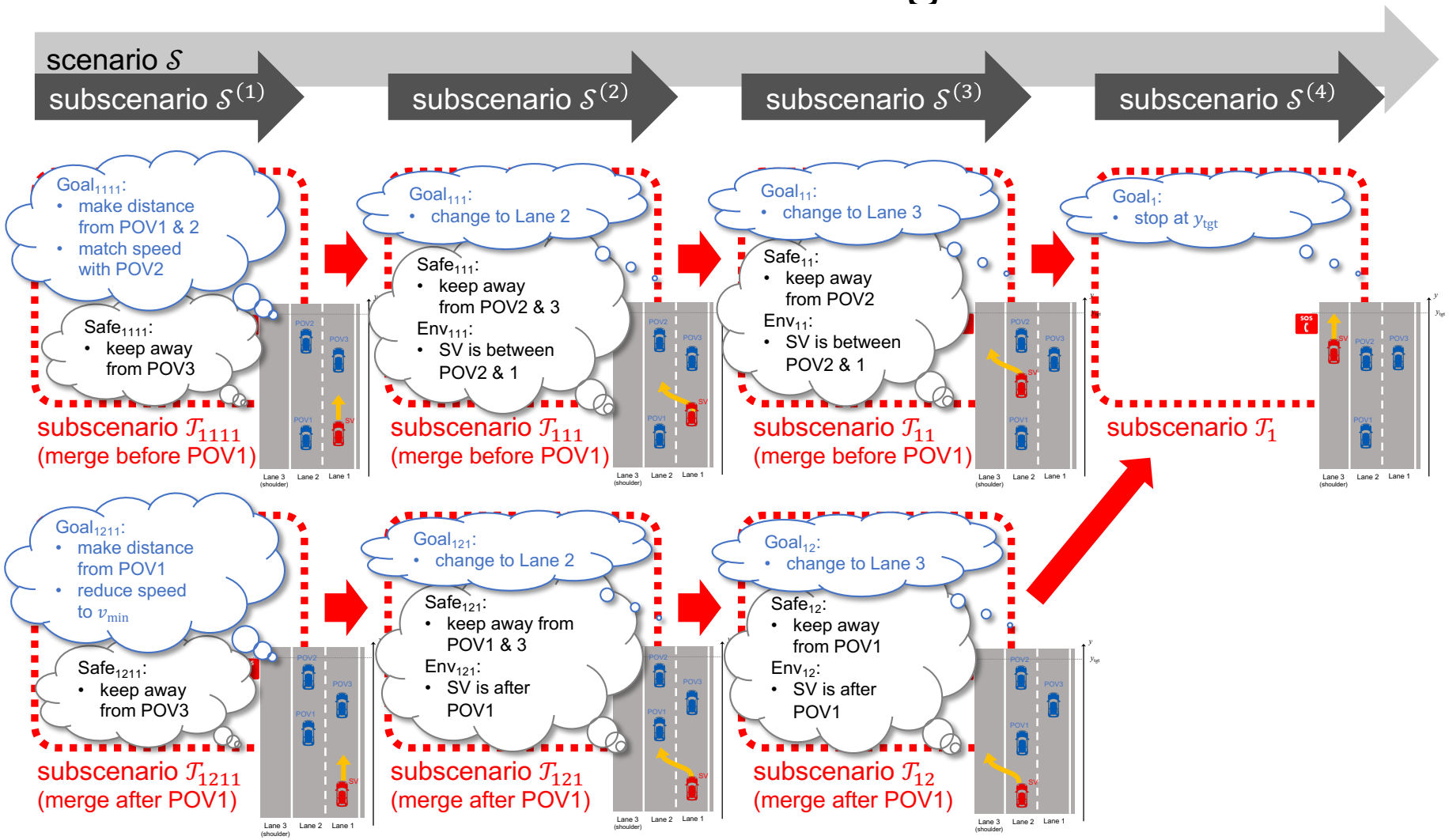
for the following given data

- $\boldsymbol{B}$ is the **goal**: "stoping on the shoulder at $\boldsymbol{y}_{\text{tgt}}$"
- $\boldsymbol{S}$ is the **safety**: "no collision," or better "securing RSS distance from every other car"

- We shall identify
  - α as an **RSS proper response**: "executing α will safely achieve the goal"
  - A as an **RSS condition**: "when A is true, B and S are guaranteed by executing α"

Compositional Rule Derivation

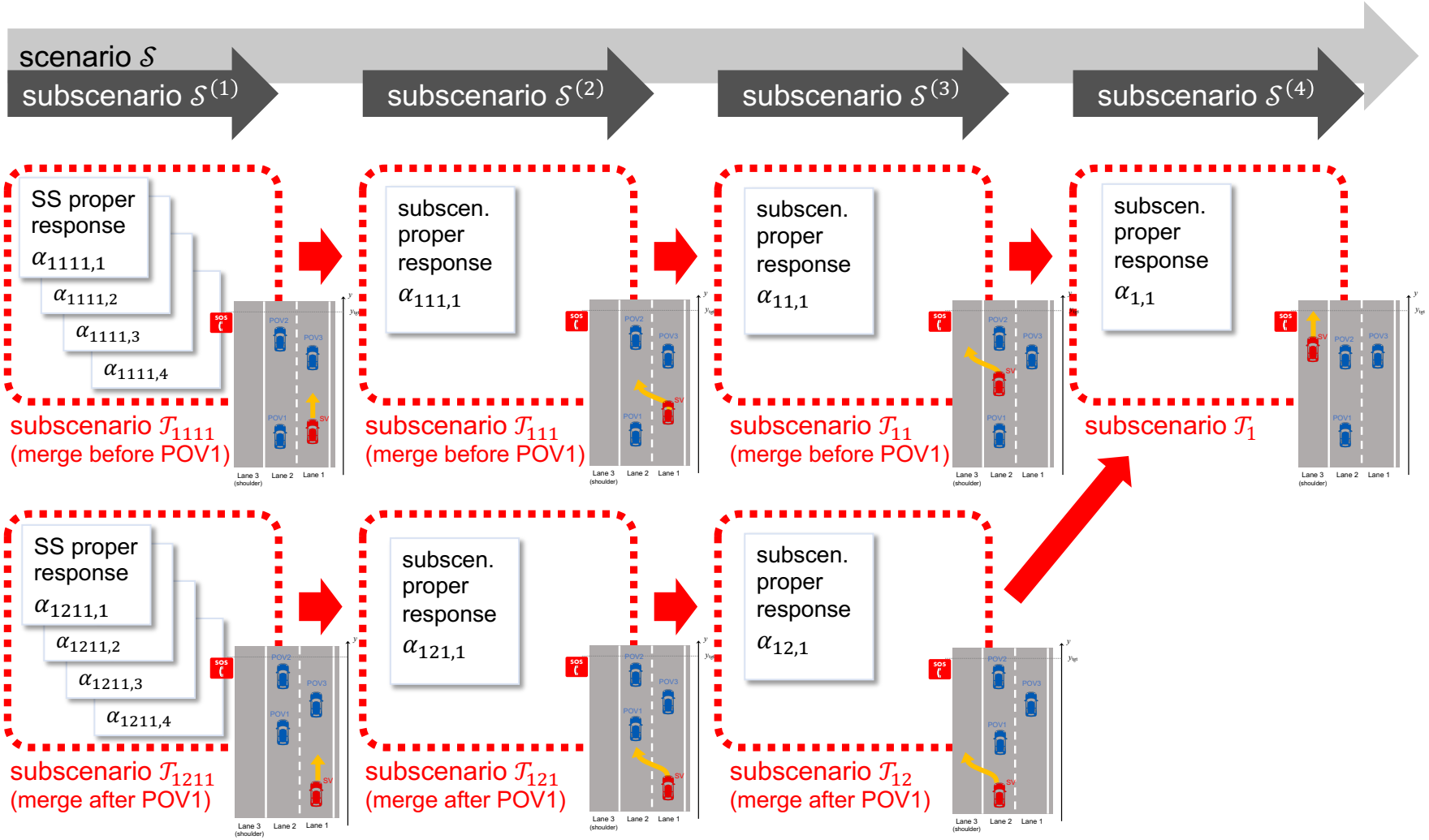
(1) Decompose the scenario into subscenarios, each of which has clearer focuses and goals



Compositional Rule Derivation

(2) Devise subscenario proper responses for each subscenario

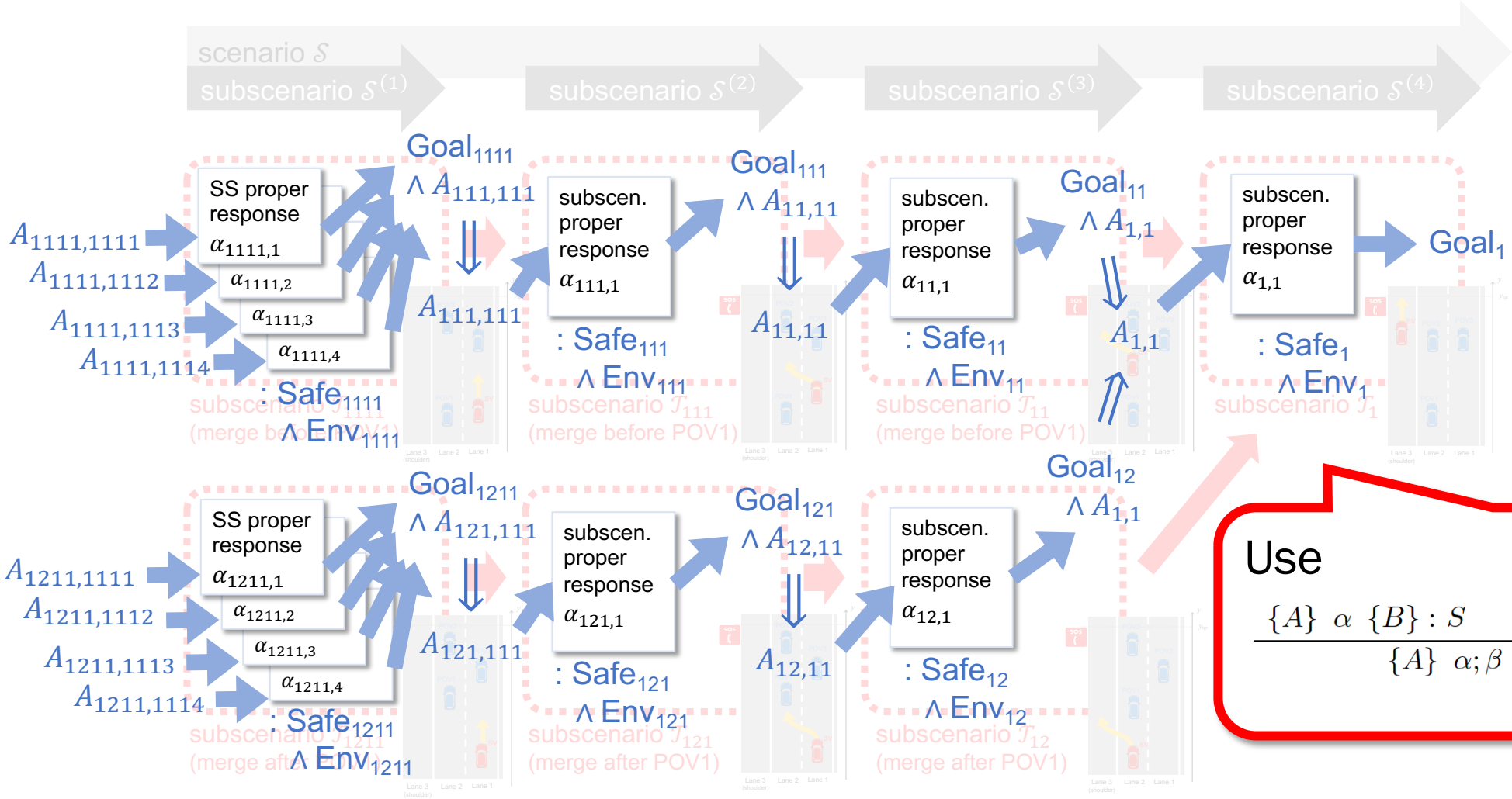
$$\{A\}\ \alpha\ \{B\} : S$$



Compositional Rule Derivation

(3) Backpropagate pre/postconditions,
leading to the scenario-wide precondition

$$\{A\}\ \alpha\ \{B\} : S$$



Use

$$\frac{\{A\}\ \alpha\ \{B\} : S \qquad \{B\}\ \beta\ \{C\} : S}{\{A\}\ \alpha;\beta\ \{C\} : S}\ (\textsc{Seq})$$

Outline

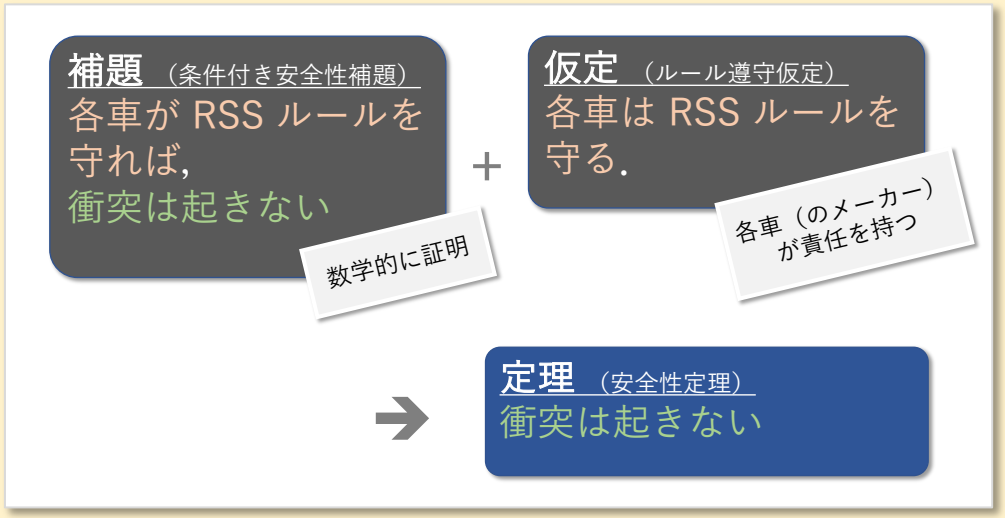
- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ

- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]

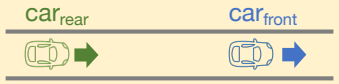
- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ

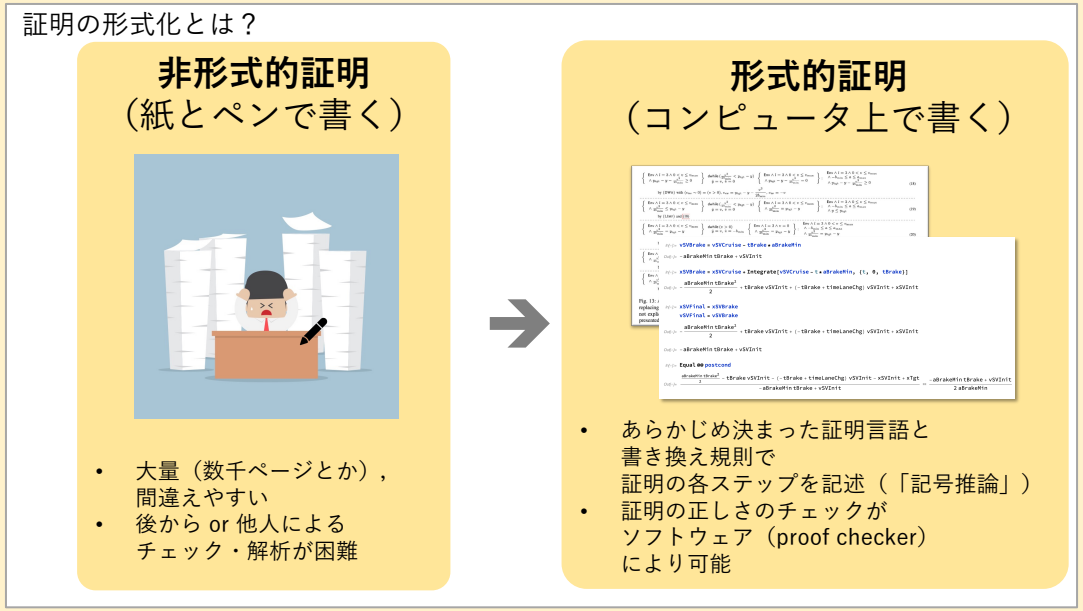
我々の成果 [Hasuo+, IEEE T-IV, 2023] の意義

論理の形式化によって RSS の適用範囲を大きく拡大
現実的運転シナリオの多くをカバー ➔ 本格的社会展開へ

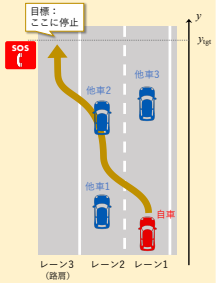


- RSS の方法論（上図 [Shalev-Shwartz et al., arXiv, 2017]）は有望だが，
  条件付き安全性補題の証明技術がなかった
- よって単純な運転シナリオに適用範囲が限定されていた





- [Hasuo+, IEEE T-IV, 2023] の貢献：
  複雑な運転シナリオに対して
  条件付き安全性補題の証明を完遂する技術
- 分割統治による証明，目標達成の保証，…
  ➔ 路肩停止など複雑なシナリオをカバー

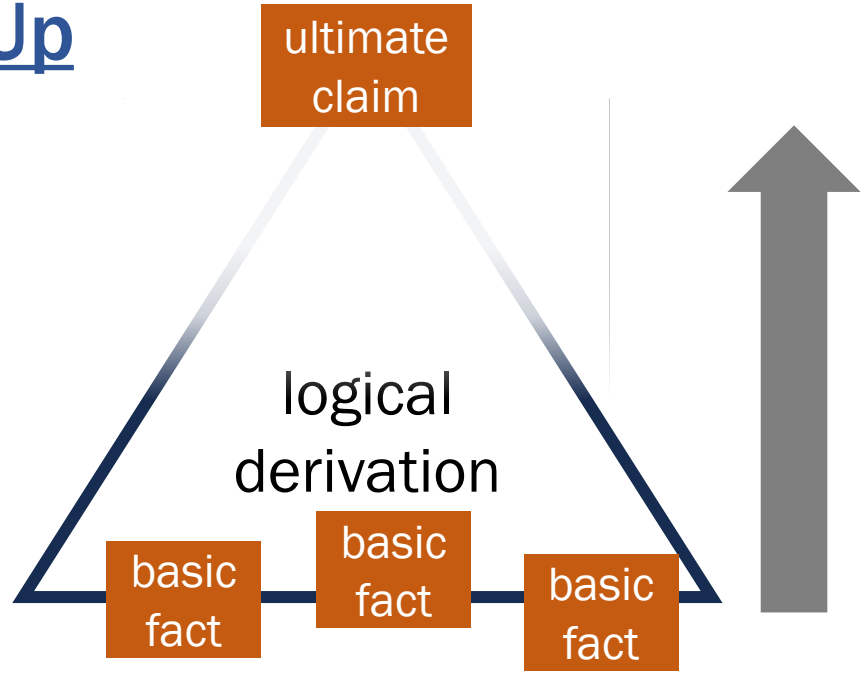


- RSS の本格展開 ➔ 自動運転の社会受容，本格普及

Use of Logic—Bottom Up or Top Down?

再掲

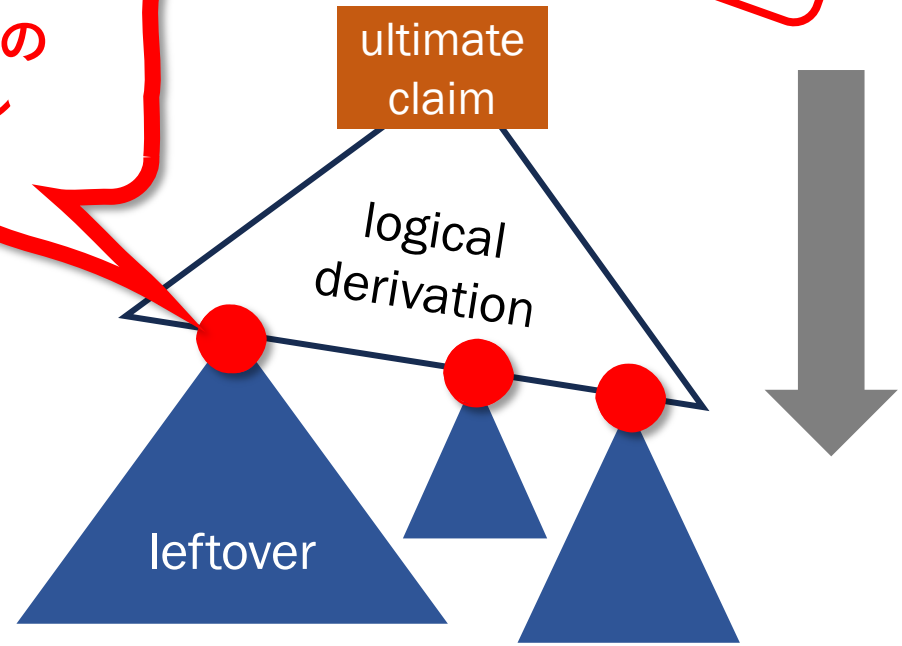
Bottom Up



Conventional use of logic

- ✓ Start from basic & unquestionable facts, and build up unquestionable facts
- ✗ Basic facts are often not available (e.g. models are rare for cyber-physical systems)
- ✗ The ultimate claim is far, far away
- ✗ Zero value in incomplete proofs

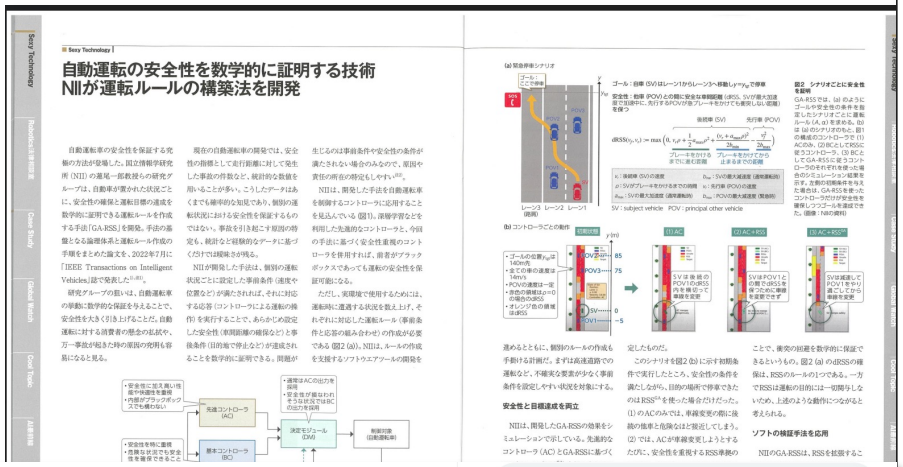
「契約」としての RSS 安全ルール



- ✗ The ultimate claim's validity is only conditional
- ✓ Scalable, best-effort (Dig as deep as the budget runs)
- ✓ Connects smoothly with statistical methods
- ✓ Handles black- & gray-box models (Test or runtime-monitor leftover assumptions)
- ✓ Logical explainability and traceability. "Logical safeguard"

メディア記事：
日経ロボティクス 2022年12月号

- 原論文の「ほぼ和訳」といった詳細な記事
- 担当ライター（今井拓司氏）と編集長（進藤智則氏）の熱意のたまもの
- 当該領域における強い関心を実感
- ホームロボット，ドローン関連からもひきあい


Sexy Technology

自動運転の安全性を数学的に証明する技術
NIIが運転ルールの構築法を開発

現在の自動運転車の開発では、安全性の指標として走行距離に対して発生した事故の件数など、統計的な数値を用いることが多い。こうしたデータはあくまでも確率的な知見であり、個別の運転状況における安全性を保証するものではない。事故を引き起こす原因の特定も、統計などで経験的なデータに基づくだけでは限界がある。

NIIが開発した手法は、個別の運転状況ごとに設定した事前条件（速度や位置など）が満たされれば、それに対応する必要（コントローラによる運転の操作）を実行することで、あらかじめ設定した安全性（車間距離の確保など）と事後条件（目的地で停止など）が達成されることを数学的に証明できる。問題が生じるのは事前条件や安全性の条件が満たされない場合のみなので、原因や責任の所在の特定もしやすい。

NIIは、開発した手法を自動運転車を制御するコントローラに応用することを見込んでいる（図1）。深層学習などを利用した先進的なコントローラと、今回の手法に基づく安全性重視のコントローラを併用すれば、搭載がブラックボックスであっても運転の安全性を保証可能になる。

ただし、実環境で使用するためには、運転時に遭遇する状況を数え上げ、それぞれに対応した運転ルール（事前条件と必要の組み合わせ）の作成が必要である（図2 (a)）。NIIは、ルールの作成を支援するソフトウエアツールの開発を


Thread

Tomonori SHINDOH
@tshindoh

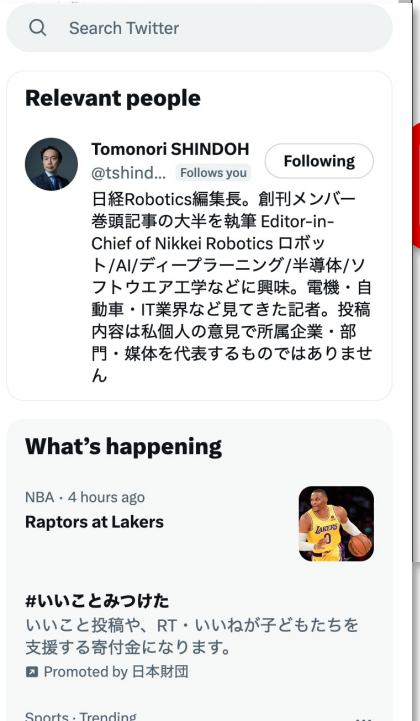
形式手法関連の記事を多く書いてきた記者として、このトピックはぜひ自媒体で掲載せねばと思ったのでした 「蓮尾先生のERATOプロジェクトはすごいですよー」と関係者の方からお聞きしていましたので -自動運転の安全性を数学的に証明する技術、NIIが運転ルールの構築法を開発

Translate Tweet

xtech.nikkei.com
《日経Robotics》自動運転の安全性を数学的に証明する技術、NIIが運転ルー...
自動運転車の安全性を保証する究極の方法が登場した。国立情報学研究所（NII）の蓮尾一郎教授らの研究グループは、自動車が置かれた状況ごとに、安...

7:22 PM · Sep 13, 2022

1 Retweet 4 Likes


Search Twitter

Relevant people

Tomonori SHINDOH
@tshind... Follows you
Following

日経Robotics編集長。創刊メンバー巻頭記事の大半を執筆 Editor-in-Chief of Nikkei Robotics ロボット/AI/ディープラーニング/半導体/ソフトウエア工学などに興味。電機・自動車・IT業界など見てきた記者。投稿内容は私個人の意見で所属企業・部門・媒体を代表するものではありません

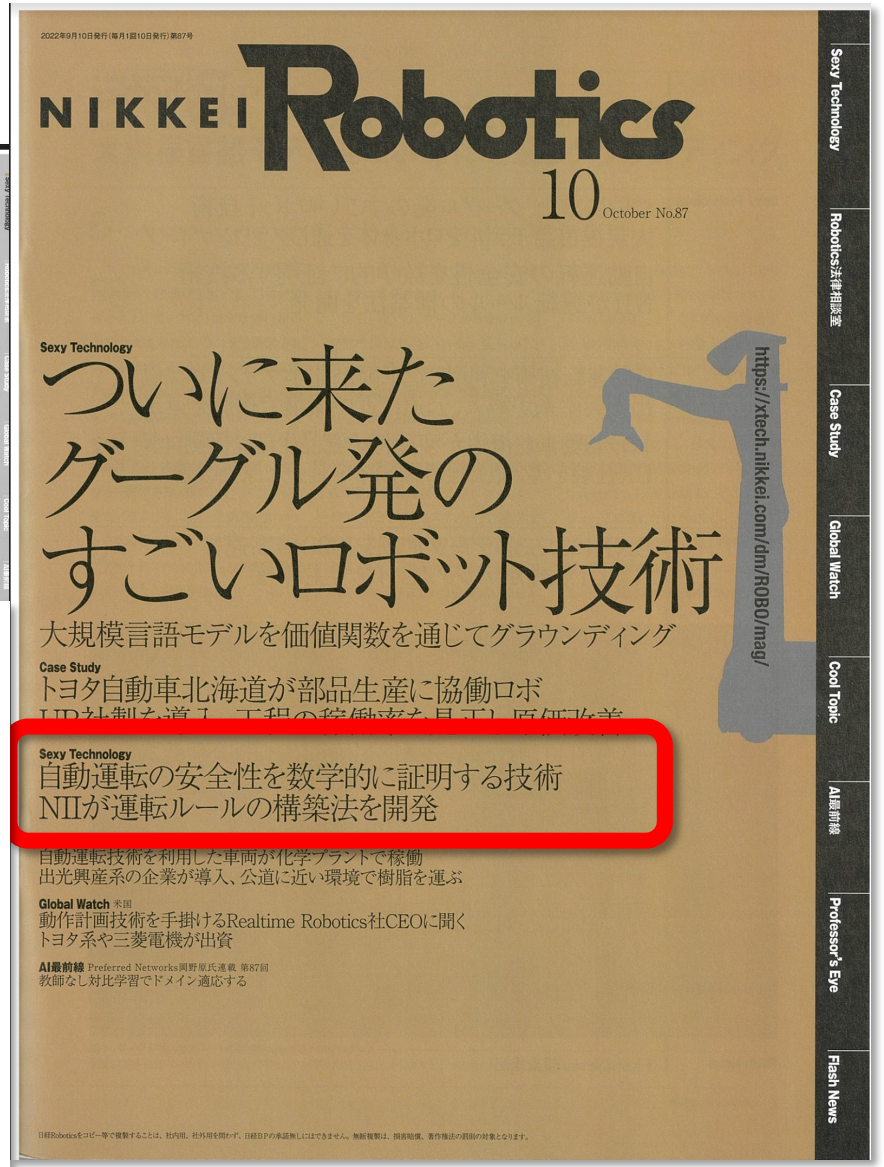
What's happening

NBA · 4 hours ago
Raptors at Lakers

#いいことみつけた
いいこと投稿や、RT・いいねが子どもたちを支援する寄付金になります。
Promoted by 日本財団

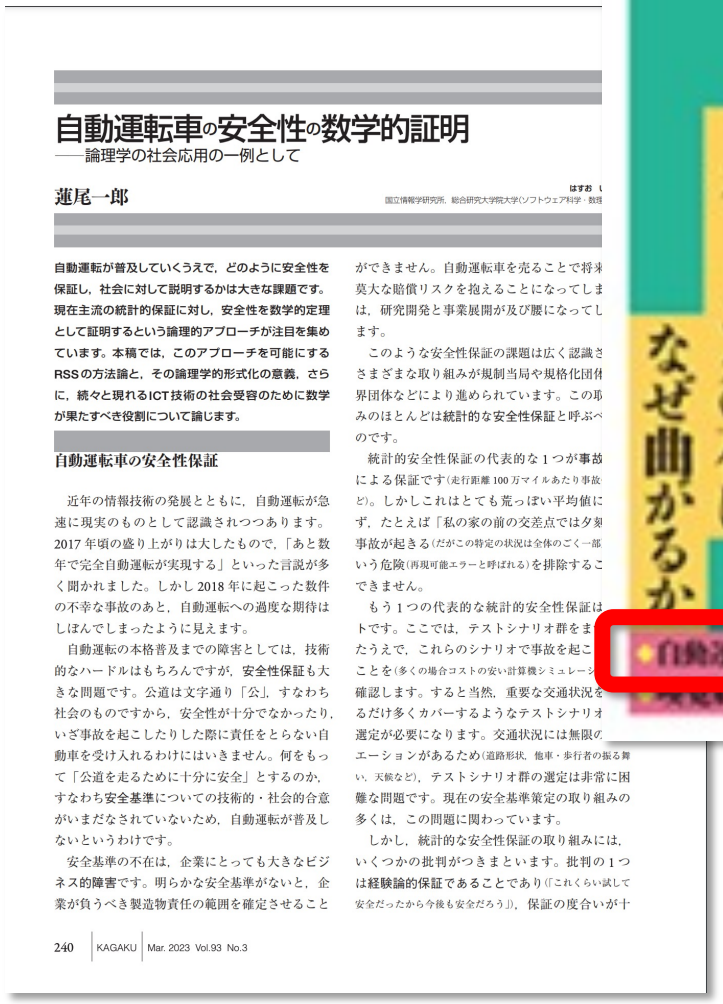
Sports · Trending
大谷翔平のただしさと息苦しさ
Trending with 佐々木朗希, チェコ戦

Sports · Trending



メディア記事 ：
岩波「科学」 2023年3月号

- 一般読者向けに蓮尾が執筆
- 自動運転
  ＋ （超高速）数理論理学入門
  ＋ 「ICT技術の説明可能性を担う
  　　社会基盤としての数理論理学」
- 数理論理学・ソフトウェア科学の成果はわかりにくいとよく言われるが,
  応用をうまく絞ると非常に関心が高い
- 自動運転本格普及の妨げとなっている
  安全性保証と社会受容の遅れに対する
  「上げ潮」的アプローチの一つ



自動運転車の安全性の数学的証明
——論理学の社会応用の一例として

蓮尾一郎

自動運転が普及していくうえで，どのように安全性を保証し，社会に対して説明するかは大きな課題です。現在主流の統計的保証に対し，安全性を数学的定理として証明するという論理的アプローチが注目を集めています。本稿では，このアプローチを可能にするRSSの方法論と，その論理学的形式化の意義，さらに，続々と現れるICT技術の社会受容のために数学が果たすべき役割について論じます。

自動運転車の安全性保証

近年の情報技術の発展とともに，自動運転が急速に現実のものとして認識されつつあります。2017年頃の盛り上がりは大したもので，「あと数年で完全自動運転が実現する」といった言説が多く聞かれました。しかし2018年に起こった数件の不幸な事故のあと，自動運転への過度な期待はしぼんでしまったように見えます。

自動運転の本格普及までの障害としては，技術的なハードルはもちろんですが，安全性保証も大きな問題です。公道は文字通り「公」，すなわち社会のものですから，安全性が十分でなかったり，いざ事故を起こしたりした際に責任をとらない自動車を受け入れるわけにはいきません。何をもって「公道を走るために十分に安全」とするのか，すなわち安全基準についての技術的・社会的合意がいまだなされていないため，自動運転が普及しないというわけです。

安全基準の不在は，企業にとっても大きなビジネス的障害です。明らかな安全基準がないと，企業が負うべき製造物責任の範囲を確定させることができません。自動運転車を売ることで将来莫大な賠償リスクを抱えることになってしまっては，研究開発と事業展開が及び腰になってしまいます。

このような安全性保証の課題は広く認識されさまざまな取り組みが規制当局や規格化団体界団体などにより進められています。この取組みのほとんどは統計的な安全性保証と呼ぶべきのです。

統計的安全性保証の代表的な1つが事故による保証です（走行距離100万マイルあたり事故など）。しかしこれはとても荒っぽい平均値にすぎず，たとえば「私の家の前の交差点では夕刻事故が起きる（だがこの特定の状況は全体のごく一部という危険（再現可能エラーと呼ばれる）を排除することできません。

もう1つの代表的な統計的安全性保証はトです。ここでは，テストシナリオ群をまたうえで，これらのシナリオで事故を起こすことを（多くの場合コストの安い計算機シミュレーション確認します。すると当然，重要な交通状況をるだけ多くカバーするようなテストシナリオ選定が必要になります。交通状況には無限のエーションがあるため（道路形状，他車・歩行者の振る舞い，天候など），テストシナリオ群の選定は非常に困難な問題です。現在の安全基準策定の取り組みの多くは，この問題に関わっています。

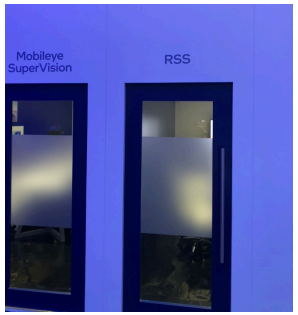
しかし，統計的な安全性保証の取り組みには，いくつかの批判がつきまといます。批判の1つは経験論的保証であることであり（「これくらい試して安全だったから今後も安全だろう」），保証の度合いが十

新たな国際展開活動を開始

国際的ニーズ・関心の高まりが 展示会，規格化WG等において可視化

我々の技術シーズへの 国際的ニーズ・関心の高まり

- CES 2023 に出席
  ➔ 自動運転の安全性保証手法への大きなニーズを実感
  （右： Mobileye 社ブース.
  定理証明手法 RSS を主要プロダクトの１つとしてフィーチャー）



- Wikipedia にも（注： 自分で書いたのではありません）

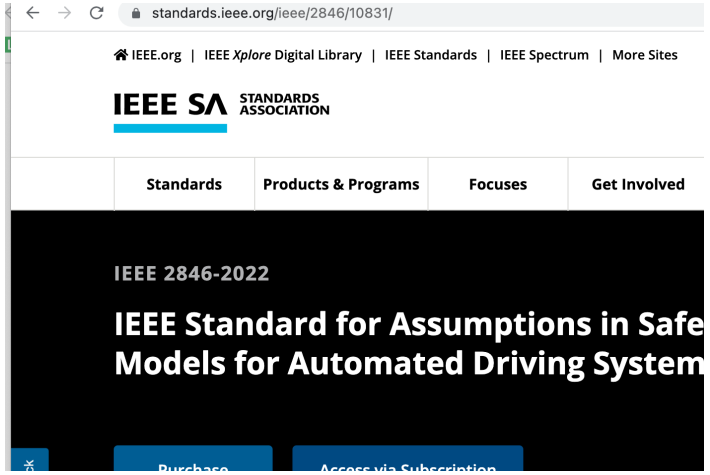
Mathematical safety model [edit]

In 2017, Mobileye published a mathematical model for automated vehicle safety which is called "Responsibility-Sensitive Safety (RSS)".[112] It is under standardization at IEEE Standards Association as "IEEE P2846: A Formal Model

In 2022, a research group of National Institute of Informatics (NII, Japan) expanded RSS and developed "Goal-Aware RSS" to make RSS rules possible to deal with complex scenarios via program logic.[114]

"Self-driving car," *Wikipedia*

国際規格化活動の本格化



- IEEE Standards Association,
  AV Decision Making WG に技術打ち込み
  ➔ 大きな関心
- Intel 社と協働
- WG メンバーとして参画（IEEE 2846）

Safety Guarantee for Automated Driving via Logical Safety Rules and Mathematical Proofs

- Decompose **safety** (a complex goal) into **logical safety rules** (explicit, easy to check and enforce)
- "Ultimate assurance" in the form of **mathematical proofs**. Logical explanation by following their reasoning steps
- Safety rules are generic and reusable
  → regulation, standard → social acceptance
- Attribution of liabilities
  (collision → someone must have broken the rules)

Regulation/Standardization Body

R_1
R_2
R_3

Safety Rule R_1

In the *same-lane same-direction* driving scenario,

- Maintain the safety distance

$$d_{\min} = \left[v_r\,\rho + \frac{1}{2} a_{\max,\mathrm{accel}}\,\rho^2 + \frac{(v_r + \rho\, a_{\max,\mathrm{accel}})^2}{2a_{\min,\mathrm{brake}}} - \frac{v_f^2}{2a_{\max,\mathrm{brake}}} \right]_+$$

  from the preceding car
- When that's hard, brake at acceleration $a_{\max,\mathrm{brake}}$

Theorem (Safety)

There is no collision attributed to the ego vehicle as long as the safety rule R_1 is respected

Proof (of the safety thm.)

⋮

The only non-obvious point is that $e_{\mathsf{inv},2}$ is preserved by the dynamics. We first observe

$$\mathcal{L}_{\delta_f,\delta_r^1} e_{\mathsf{inv},2} = \begin{cases} 0 & \text{if } \mathsf{dRSS}_{\pm}(v_f, v_r, \rho - t) \geq 0 \\ v_f - v_r & \text{otherwise,} \end{cases}$$

where $\mathsf{dRSS}_{\pm}(v_f, v_r, \rho)$ is given by

$$\mathsf{dRSS}_{\pm}(v_f, v_r, \rho) = v_r\rho + \frac{a_{\max}\rho^2}{2} + \frac{(v_r + a_{\max}\rho)^2}{2b_{\min}} - \frac{v_f^2}{2b_{\max}}.$$

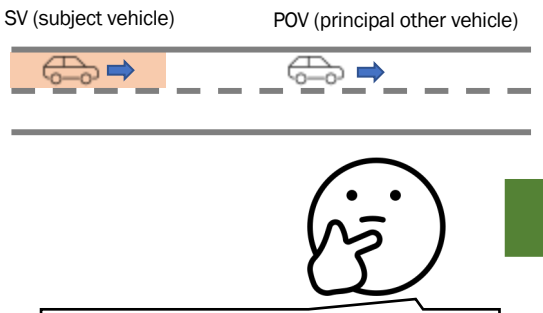
Therefore, we can infer as follows.

$$\mathsf{dRSS}_{\pm}(v_f, v_r, \rho - t) < 0$$
$$\iff v_r(\rho - t) + \frac{a_{\max}(\rho - t)^2}{2} + \frac{(v_r + a_{\max}(\rho - t))^2}{2b_{\min}} - \frac{v_f^2}{2b_{\max}} < 0$$

⋮

RSS 技術の応用

社会的契約として RSS ルールを活用
自動運転エコシステムのあらゆる場面で大きなインパクト



「現時点でこのルール C を守っていれば，先行車が急ブレーキをかけても安全に停止できる」…というような安全ルール C を求めたい

ルール導出技術者が各運転シナリオに対し個別にルール導出

- 同時に，ルールCの安全性証明も書き出す

安全ルール C

＋

安全ルールCの安全性証明

規制当局他，誰もが正当性をチェックできる

- 事故の責任所在の特定
  ルールCを破ったあなたが悪い
- 自動運転の安全性規準・規格
  （ルール遵守を認証しないと販売不可）
  IEEE P2846 など

受益者：規制当局，規格化団体，業界団体

- 自動運転車の安全性証明・アピール ＆ 製造物責任の明示化・限定
  ルール C の遵守を示せば十分
- 自動運転車の実行時監視
  ルール C を破りそうになったら安全系が介入
  → 安全性保証，通常系設計の自由度向上

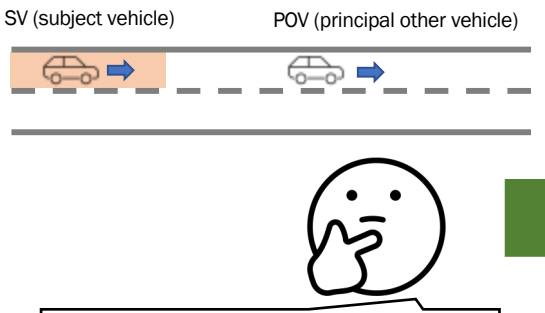
受益者：自動運転システムベンダー

- 保険料率計算
  ルール C を守っているので保険料を安く

受益者：損害保険

RSS 技術の応用

社会的契約として RSS ルールを活用
自動運転エコシステムのあらゆる場面で大きなインパクト



「現時点でこのルール C を守っていれば，先行車が急ブレーキをかけても安全に停止できる」…というような安全ルール C を求めたい

ルール導出技術者が各運転シナリオに対し個別にルール導出

- 同時に，ルールCの安全性証明も書き出す

安全ルール C

＋

安全ルールCの安全性証明

規制当局他，誰もが正当性をチェックできる

事故の責任所在の特定
ルールCを破ったあなたが悪い

自動運転の安全性規準・規格
（ルール遵守を認証しないと販売不可）
IEEE P2846 など

受益者：規制当局 規格化団体 業界団体

自動運転車の安全性証明・アピール
＆ 製造物責任の明示化・限定
ルール C の遵守を示せば十分

自動運転車の実行時監視
ルール C を破りそうになったら安全系が介入
→ 安全性保証，通常系設計の自由度向上

受益者：自動運転システムベンダー

保険料率計算
ルール C を守っているので保険料を安く

受益者：損害保険

RSS ルールによる安全エンベロープ

既存のコントローラに冗長系・安全系として付加可能
コントローラを監視，必要になれば介入 ➔ 安全性保証

RSS ルールの例

一車線同方向運転シナリオにおける追突回避のためのルール

[Shalev-Shwartz et al., arXiv preprint, 2017]

RSS 条件：
車間距離を

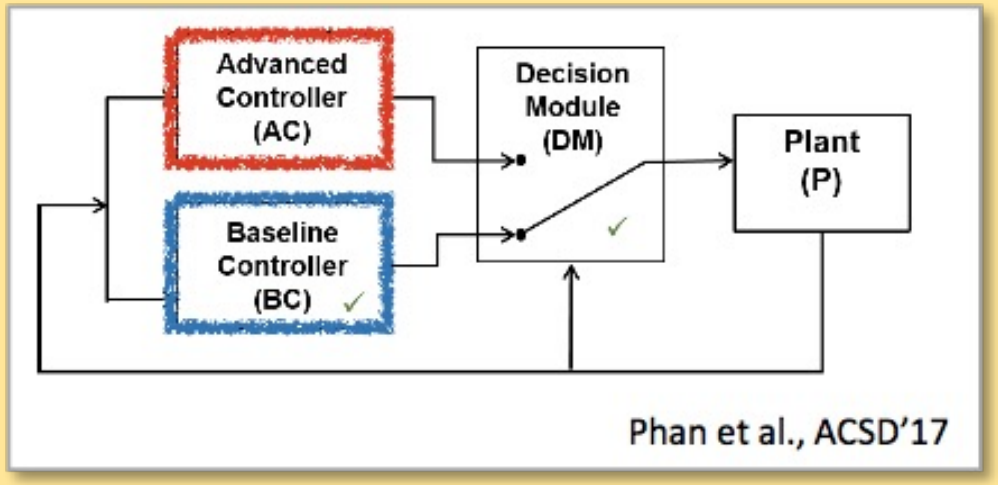
$$d_{\min} = \left[v_r \rho + \frac{1}{2} a_{\max,\text{accel}} \rho^2 + \frac{(v_r + \rho a_{\max,\text{accel}})^2}{2a_{\min,\text{brake}}} - \frac{v_f^2}{2a_{\max,\text{brake}}} \right]_+$$

以上確保すること

適切反応 (proper response)：
上記RSS 条件の違反が予見される場合，
反応時間 ρ 以内に 減速度 $a_{\min,\text{brake}}$ でブレーキすること

条件付き安全性補題：
RSS 条件が真である状態から適切反応を実行すれば，衝突は発生しない

- RSS ルールの構造
  - RSS 条件：
    「この条件を満たしているうちは
    まだ逃げ道がある」
  - 適切反応：
    「やばくなったらこうして逃げろ」

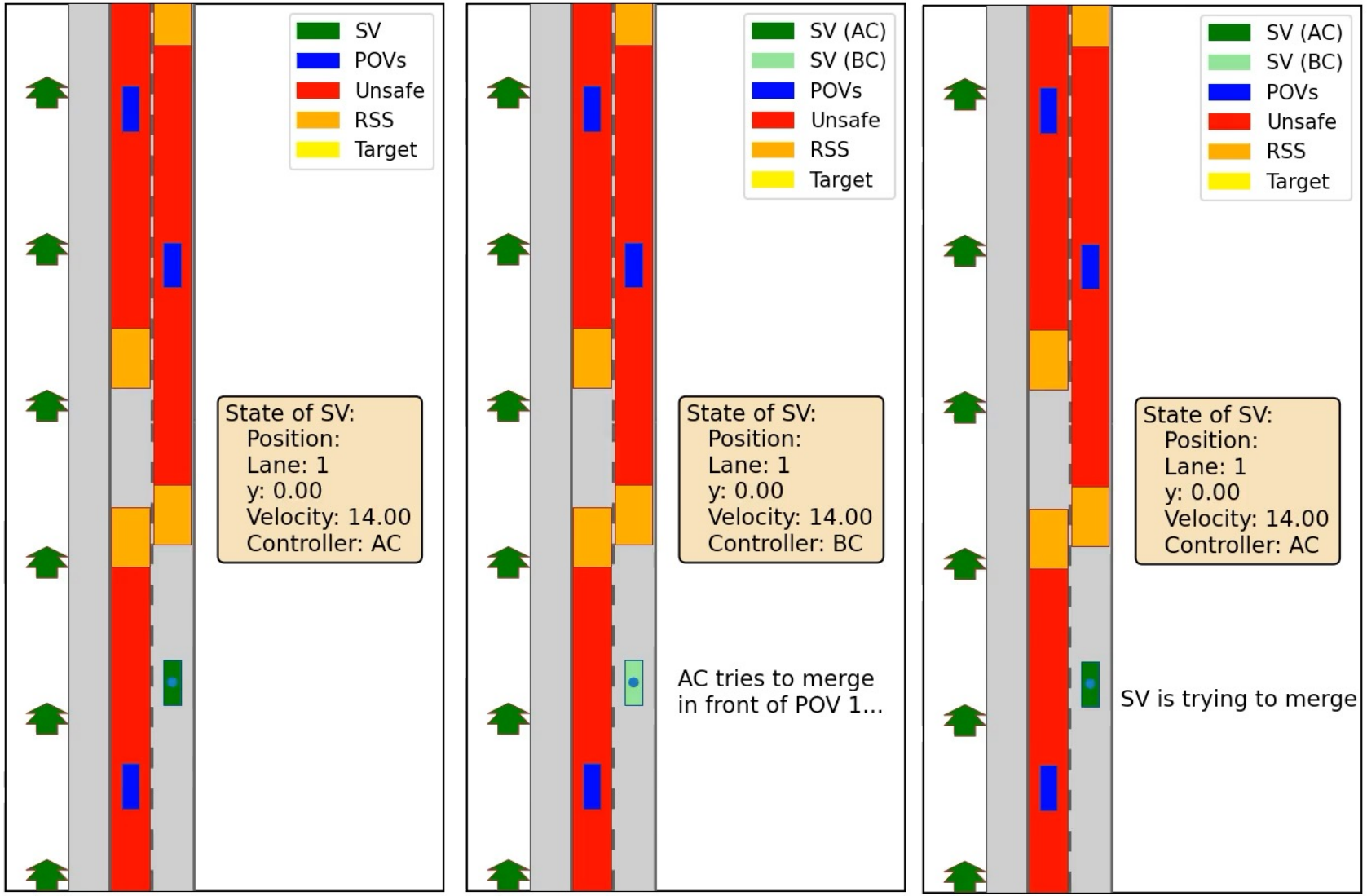


- 安全アーキテクチャ
  - AC（通常系）が性能を追求
  - BC（安全系）は安全性のみを追求
  - Decision module（DM）が両者を切り替え
    （危なくなったら BC を使う）
- RSS ルールがぴったりはまる
  - AC： 既存のコントローラ
  - BC： 適切反応を実行
  - DM： RSS 条件が成立するか監視.
    成立が怪しくなってきたら AC ➔ BC 切替え

安全エンベロープ実行例１

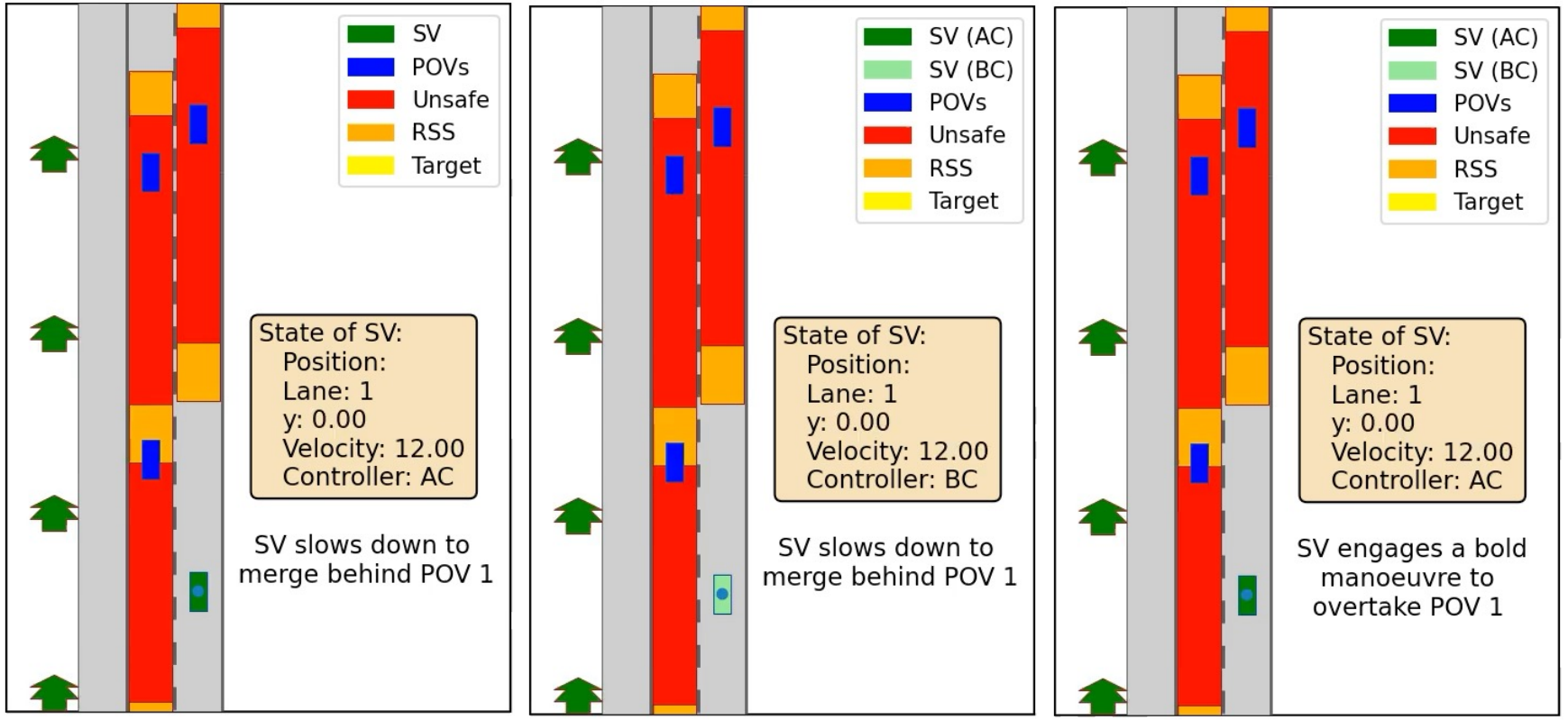
- AC: 安全系なし
- AC+RSS：
  既存研究 [Shalev-Shwartz et al., arXiv, 2017] の RSS ルールを安全エンベロープとして実装
  （近視眼的に衝突回避）
- AC+RSS^{GA}：
  我々 [Hasuo+, IEEE T-IV] の RSS ルールを安全エンベロープとして実装
  （長期的視野で目標達成も保証）

- AC は安全でない（危険な割り込み）
- AC+RSS は路肩に到達できず
- AC+RSS^{GA} は長期的視野で減速 ➔ 他車の後ろで合流，安全性と目標達成（路肩停止）両方を実現



安全エンベロープ実行例２

- AC: 安全系なし
- AC+RSS：
  既存研究 [Shalev-Shwartz et al., arXiv, 2017] の RSS ルールを安全エンベロープとして実装
  （近視眼的に衝突回避）
- AC+RSS^{GA}：
  我々 [Hasuo+, IEEE T-IV] の RSS ルールを安全エンベロープとして実装
  （長期的視野で目標達成も保証）

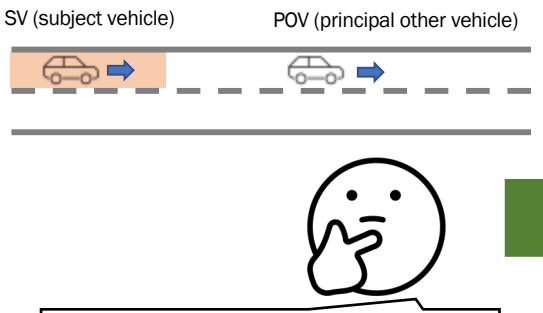


AC　　AC+*RSS*　　AC+*RSS*^{GA}

- AC, AC+RSS は安全な路肩停止に成功，しかし遅い
- AC+RSS^{GA} は（安全性保証のもと）加速して追い越し，
  他車の前で合流して路肩停止
- ➔「自動運転車は安全性ばかり気にして前に進まない」へのアンチテーゼ！

再掲

社会的契約として RSS ルールを活用
自動運転エコシステムのあらゆる場面で大きなインパクト



「現時点でこのルール C を守っていれば，先行車が急ブレーキをかけても安全に停止できる」…というような安全ルール C を求めたい

ルール導出技術者が各運転シナリオに対し個別にルール導出

- 同時に，ルールCの安全性証明も書き出す

安全ルール C

＋

安全ルールCの安全性証明

規制当局他，誰もが正当性をチェックできる

- 事故の責任所在の特定
  ルールCを破ったあなたが悪い
- 自動運転の安全性規準・規格
  （ルール遵守を認証しないと販売不可）
  IEEE P2846 など

受益者：規制当局，規格化団体，業界団体

- 自動運転車の安全性証明・アピール ＆ 製造物責任の明示化・限定
  ルール C の遵守を示せば十分
- 自動運転車の実行時監視
  ルール C を破りそうになったら安全系が介入
  → 安全性保証，通常系設計の自由度向上

受益者：自動運転システムベンダー

- 保険料率計算
  ルール C を守っているので保険料を安く

受益者：損害保険

自動運転技術の本格普及へ

２つのビジネス展開，異なるアプローチ

固定ルートバス・タクシー・運送

自家用車のベストエフォート自動運転

リモート監視	人間による介入	ドライバーが乗車
固定ルートのモビリティ・運送サービス	ビジネス	自動運転機能を持つ自家用車を販売
あり （既定のルートのみ運行）	ジオフェンス	なし （公道すべての走行を想定）
フルODD （ルート全てでの自動運転を想定）	ODD （operational design domain, どの状況で自動運転するか？）	部分的ODD （自動運転を行う状況を事前に規定 例： 高速道路）

自動運転技術の本格普及へ

２つのビジネス展開，異なるアプローチ

どちらにしろ，責任をとるためには，
運転シナリオを事前に知る必要がある

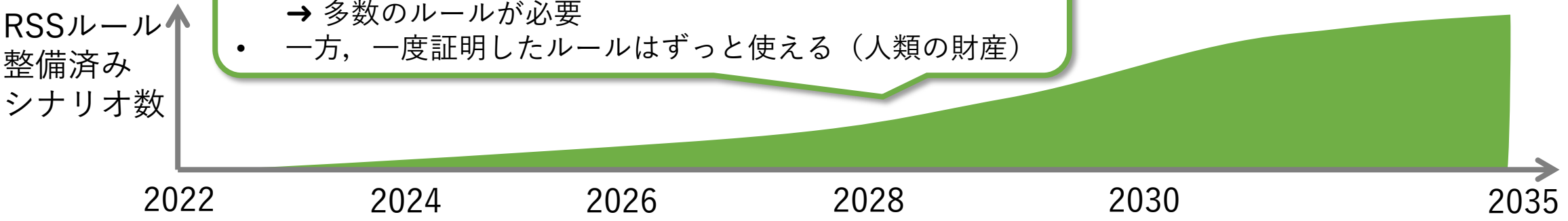
→ これら運転シナリオに対し
RSS ルールを事前策定，
数学的証明付きの安全性を実現

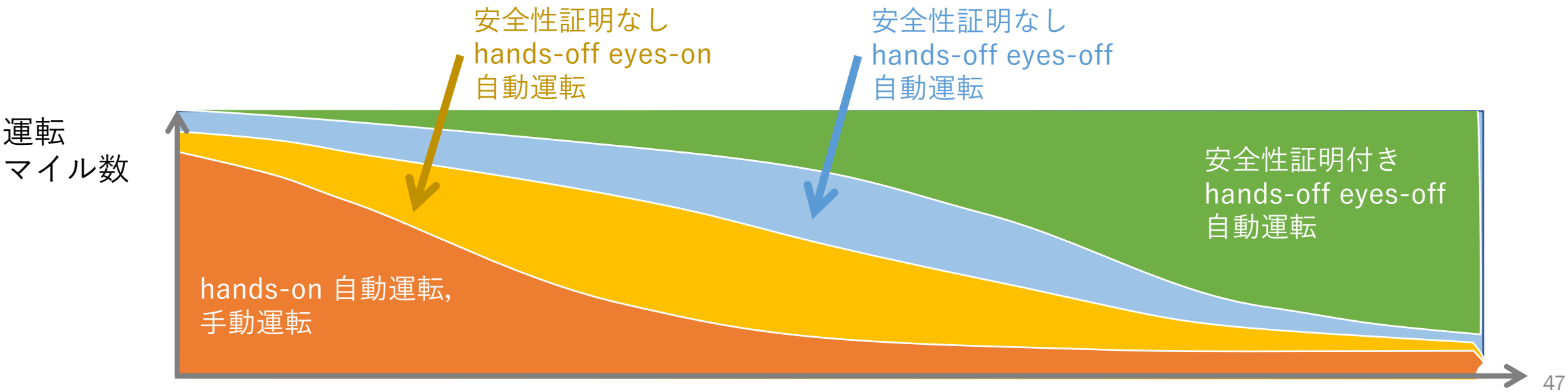
固定ルートバス・タクシー・		ベストエフォート自動運転
リモート監視	人間による介入	ドライバーが乗車
固定ルートのモビリティ・運送サービス	ビジネス	自動運転機能を持つ自家用車を販売
あり （既定のルートのみ運行）	ジオフェンス	なし （公道すべての走行を想定）
フルODD （ルート全てでの自動運転を想定）	ODD （operational design domain, どの状況で自動運転するか？）	部分的ODD （自動運転を行う状況を事前に規定 例： 高速道路）

自動運転技術の本格普及へ

RSS ルールの漸進的整備と歩みをあわせた ODD の漸進的拡大

- １シナリオあたり１ヶ月・人程度の工数を想定
- 法規・慣習が違うとルールも違う
  ➔ 多数のルールが必要
- 一方，一度証明したルールはずっと使える（人類の財産）





Outline

- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ
- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]
- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ

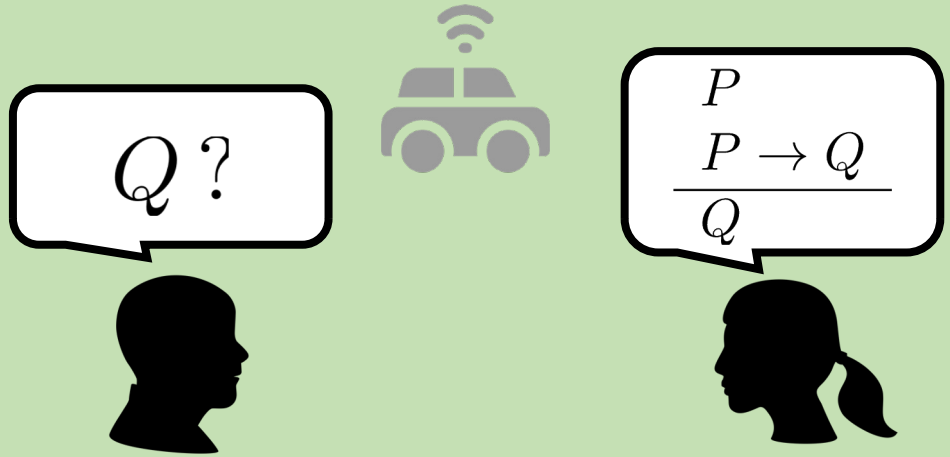
自動運転車の安全性保証： 2つの未来像

ブラックボックス 安全性保証

- 中身の見えない「全体主義的」安全性論証
- 議論，精査，批判的検討，改善が困難

vs

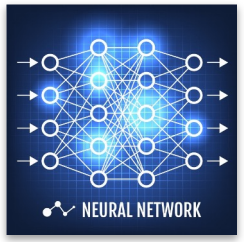
説明可能な 安全性保証



- 説明責任，トレーサビリティを満たす安全性論証を論理的に構成
- 自動運転車の安全性保証は終わりがない課題
  ➔ 社会全体での取り組みをサポート
- 我々の目指す未来像

自動運転車の安全性のための論理的技術で急速に発展する統計的技術を補完

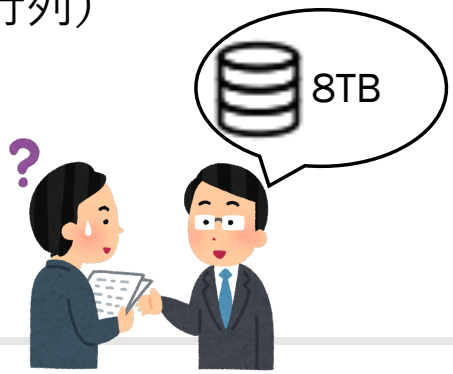
我々の技術

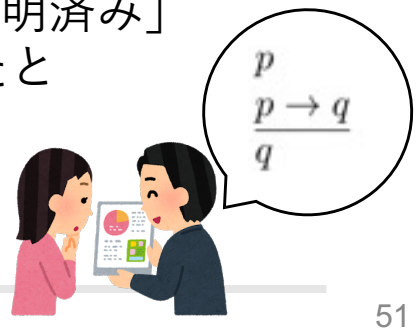
統計的手法 			論理的手法 $\dfrac{P \quad P \to Q}{Q}$	
ビッグデータから「学習」した 巨大行列，数値的		意思決定のルール	人間に理解可能な論理的ルール， 記号的	
✓	数値的正規化 入力が大体同じ → 出力もだいたい同じ	ノイズ耐性	✗ → ✓	厳格すぎるルール適用 → 実行時監視でノイズ許容
✓	ビッグデータから数値的最適化で ルールを自動学習	スケーラビリティ 大規模化	✗ → ✓	論理的ルールの設定と精査 は人力，工数大 → 証明自動化で工数削減
✗	正しさの保証なし， 数値的ルールの内容の理解は 人間には困難	説明可能性	✓	論理的ルールは理解可能， ルールの適用による推論も 論理的に追跡可能

社会的説明 ＝ 人間のコミュニケーションは ビッグデータでなく論理

我々の技術

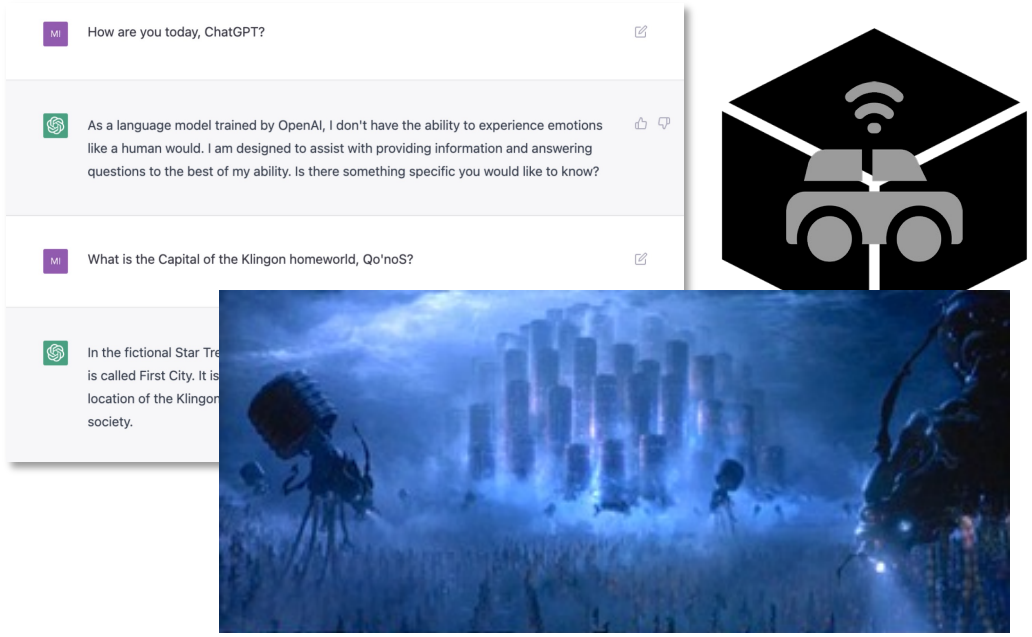
統計的手法		論理的手法 $\dfrac{P \quad P \to Q}{Q}$
「なぜこの写真をみて赤信号だと考えたのか？」 → 「NN の重み行列がコレ（10000 x 10000 行列）だったから」 「NN をこのデータ（8TB）で訓練したから」	「なぜこう考えたのか？」の説明	「なぜ前の車を追い越せると考えたのか？」 → 「このルール（数学的条件）に従った」 「このルールに従えば安全であると，仮定 A, B, C のもとで証明済み」 「よって今回危険であったとすれば，仮定 A, B, C のどれかが成り立たなかったということ」

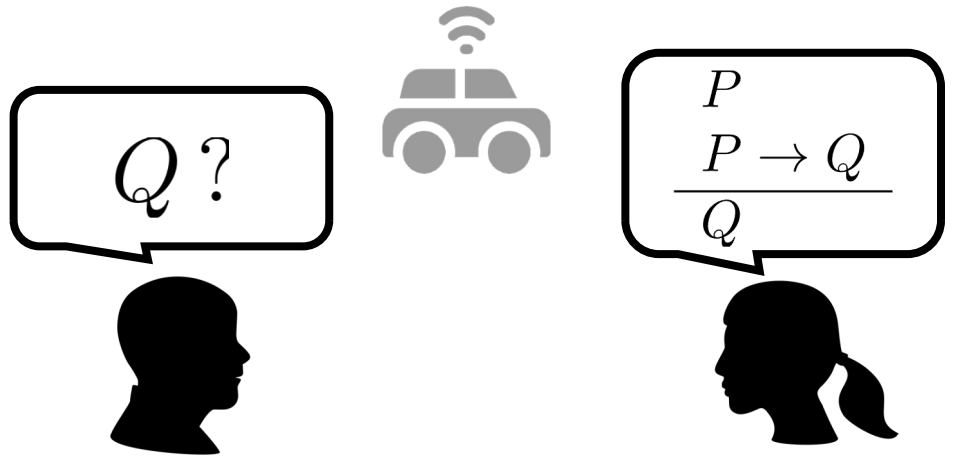




数学・論理学の社会的使命

人命に関わる技術はブラックボックスであってはいけない
証明による仮定・契約の明示化，責任所在の明確化





- ブラックボックスの ICT 新技術は safety-critical 応用に普及しないし，普及させてはいけない
- 絶対の真理（のみ）を導く手段としての証明から，仮定・契約を明示化して，責任所在を明確にし，安全性の議論の拠り所となる
  コミュニケーションメディアとしての証明，
  社会インフラとしてのソフトウェア科学

The Modeling Problem in Emerging ICT

- Theorems need *definitions*;
  formal verification needs *modeling*

再掲
「来るべきICT技術」における方法論的困難

Emerging ICT

Conventional ICT Systems

- They operate following a logical recipe (= program)
- Programs are mathematical models

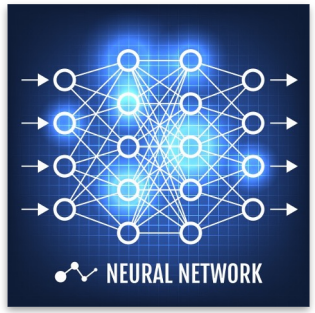
✓

Cyber-Physical Systems

- E.g. automated driving cars
- Physical components? Other cars? Pedestrians?

??

Statistical AI Systems



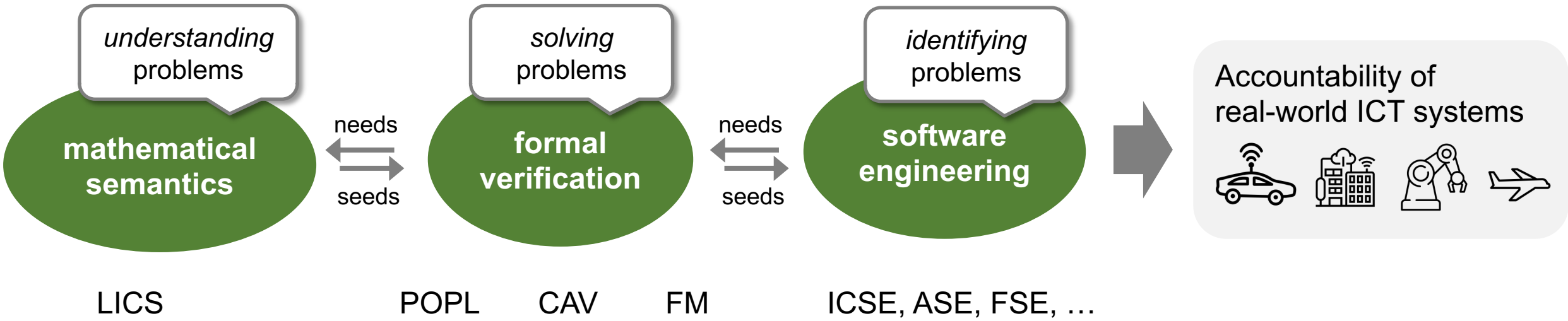
- They decide using matrices learned from big data
- Those matrices are too big to logically analyze

??

...

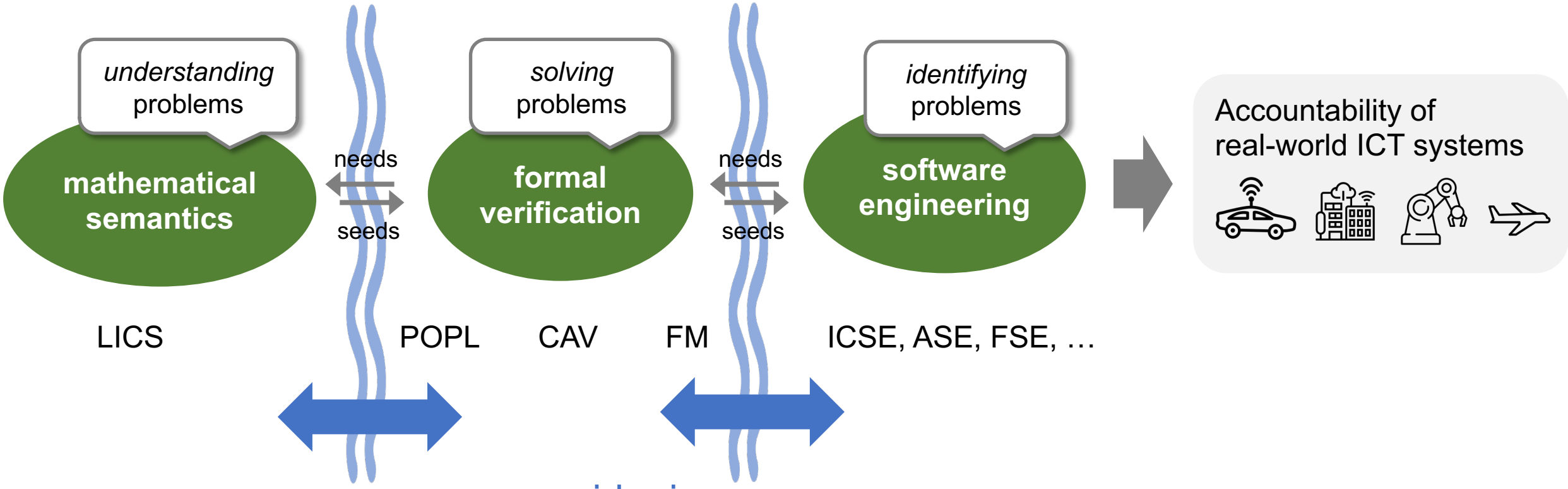
Software Science for Accountable ICT

Three Communities



LICS POPL CAV FM ICSE, ASE, FSE, …

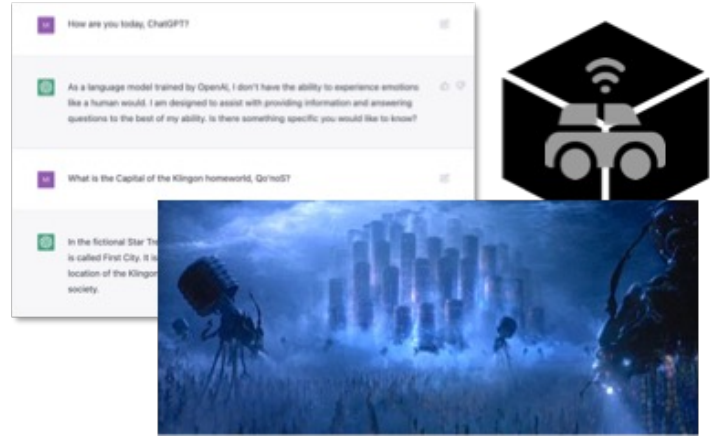
Centrifugal Developments, Widening Gaps

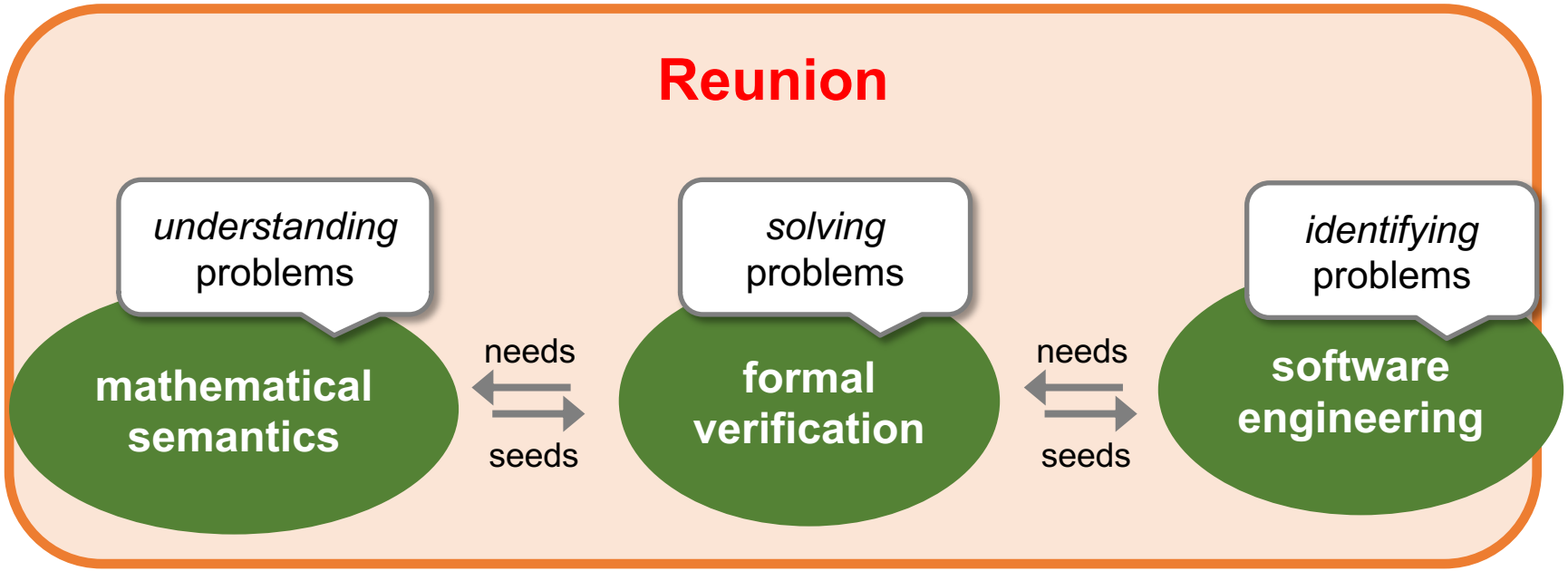


- 30 years ago, big CS labs used to accommodated both theoretical and practical research
- Now this is very rare

Software Science for Accountable ICT

Time to Reunite





Accountability of real-world ICT systems

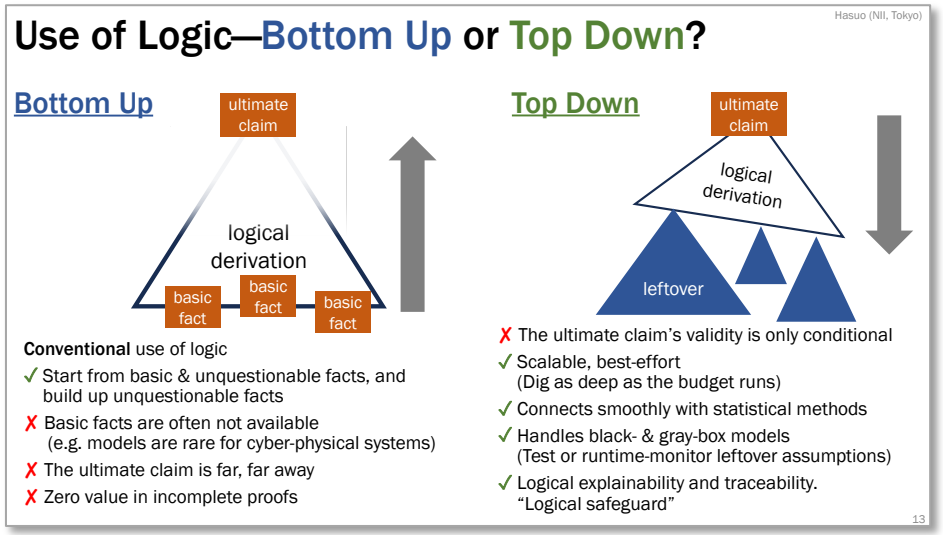
- Good for application
- Good for theory, too

Our preliminary results:

- automated driving
- compositional model checking

Summary

- ソフトウェア科学，論理学，数学的証明
  - 証明には定義が必要 ➔ モデリングの課題
  - 論理学の使い方： トップダウン，ボトムアップ



- 研究成果： 自動運転車の安全性の数学的証明
  [Hasuo et al., IEEE Trans. Intell. Vehicles, 2023]

- 来るべき情報技術の社会的信頼樹立に向けて
  - 「自動運転車安全性証明」の成果の社会展開
  - 数学的証明・ソフトウェア科学の社会的役割
  - ソフトウェア科学の再結集へ