Consensus Building in Level 4 Automated Driving Field Trials through Assurance Cases

Yutaka Matsuno
Nihon University, Japan
matsuno.yutaka@nihon-u.ac.jp

Michio Hayashi and Tomoyuki Tsuchiya
TIER IV, North America & Japan
{michio.hayashi.2,tomoyuki.tsuchiya}@tier4.jp

This research is based on results obtained from a project, JPNP21027,subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

Contents

Background

- SAE Level 4 (L4) automated driving systems are "open systems"
- The environment continuously evolves and uncertainties increase
- Openness broadens the group of stakeholders that are part of the system
  - *Internal stakeholders*: CxO, Fellow, Architect, Business, R&D,…
  - *External stakeholders*: Citizens, City Officials, Police, Nation, Investor,…

Consensus Building among Stakeholders
(safety expert/non-expert)

TIER IV L4 Automated Driving Demonstration

- TIER IV, an automated driving startup, began planning an SAE L4 demonstration in a city in Japan, Nov 2024
  - Successfully conducted without any incidents in Jan 2025
- We detail how the demonstration was planned, prepared, and conducted, focusing on consensus building regarding safety among internal stakeholders

Related Work (1/2)

Assurance Cases for AD Systems

- Patterns
- Standards
  - ISO 26262, SOTIF, UL 4600

Confidence Assessment Methods (CAMs)

- Expert-based scoring
- Probabilistic models
- Eliminative argumentation
- Bayesian networks

Related Work (2/2)

Safety Communication Practice

- UL 4600: Engineer-centric detailed template
- SAFAD: 12 safety principles with V&V roadmap
- NHTSA VSSA: Public-facing safety booklets

Our Activities in Japan

- Assurance Cases and GSN adoption since 2009
- Focusing on Consensus Building

Challenges

- GSN Communication: Difficult to effectively communicate with safety non-experts
- One-way communication
- Consensus Assessment: Lack of means to measure agreement across diverse stakeholders

Our Contributions

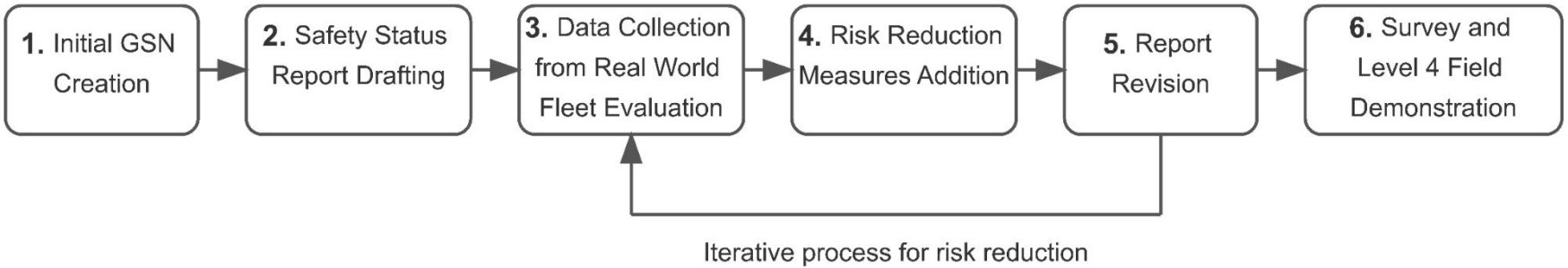
- Stakeholder-oriented safety communication framework
  - Safety Status Report: Plain-language complement to GSN-based arguments
  - Two-way communication by questionnaire
- Consensus Score

L4 Demonstration Context

- *Location:* A Japanese City
- *Vehicle:* Minibus (BYD J6)
- *Planning*: Nov. 2024
  > *Launch:* Jan. 2025
- *Challenge:* No formal safety report initially

Process Overview

1. Initial GSN Creation
2. Safety Status Report Drafting with natural language
3. Real-world data collection
4. SOTIF Alignment Loop
5. Questionnaire Survey
6. L4 Demonstration



Initial GSN Development

- Attempted GSN-based assurance cases using existing development artifacts
- *Result:* Multiple defeaters and insufficient evidence to justify safety claims
- *Decision:* Document current limitations rather than complete assurance

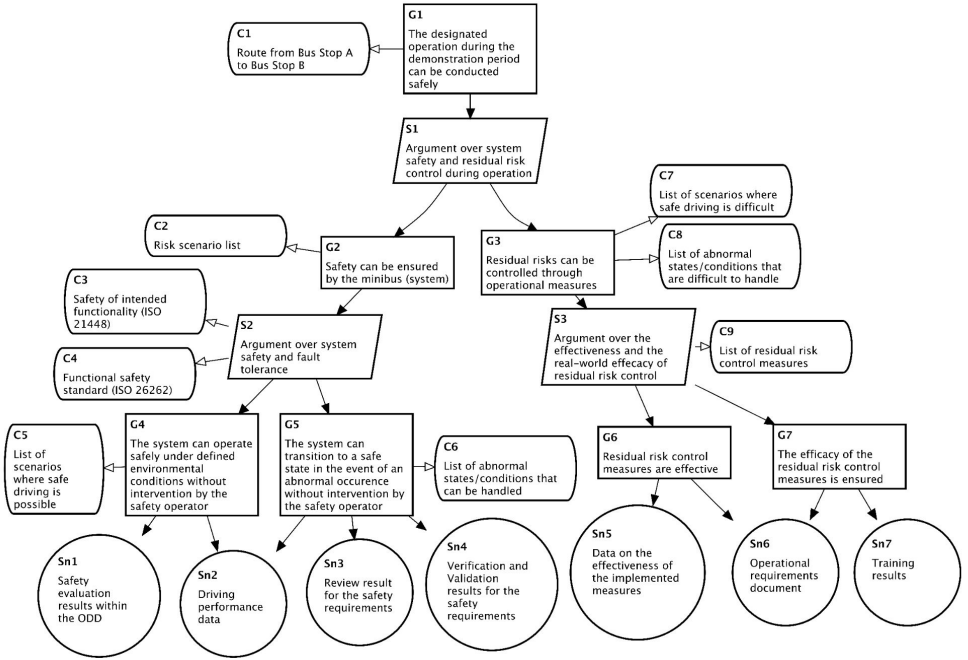
Our Approach

- Transparently communicate current safety status, not to claim complete safety
  (≠ Safety Case Report)
- Continuously updated with fleet evaluation data

Safety Status Report (SSR)

GSN Model Structure

- Approach
  - Deductive argument from inductive analysis of existing artifacts
- Design Choice
  - Abstraction level chosen to facilitate stakeholder discussions
  - Balances technical detail with accessibility for non-safety experts

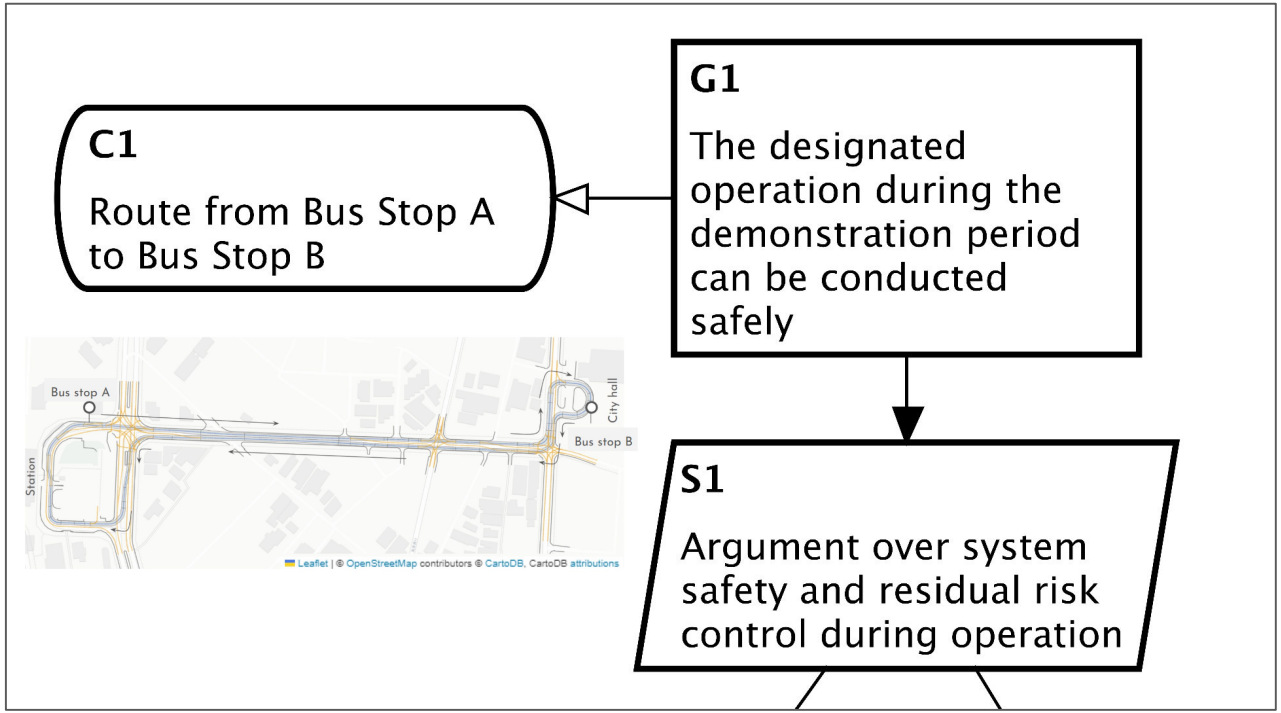


GSN Model Top Structure

G1: Designated operation during demonstration

S1:System safety and operational risk decomposition

Operation scope: Automated minibus from Bus Stop A to Bus Stop B



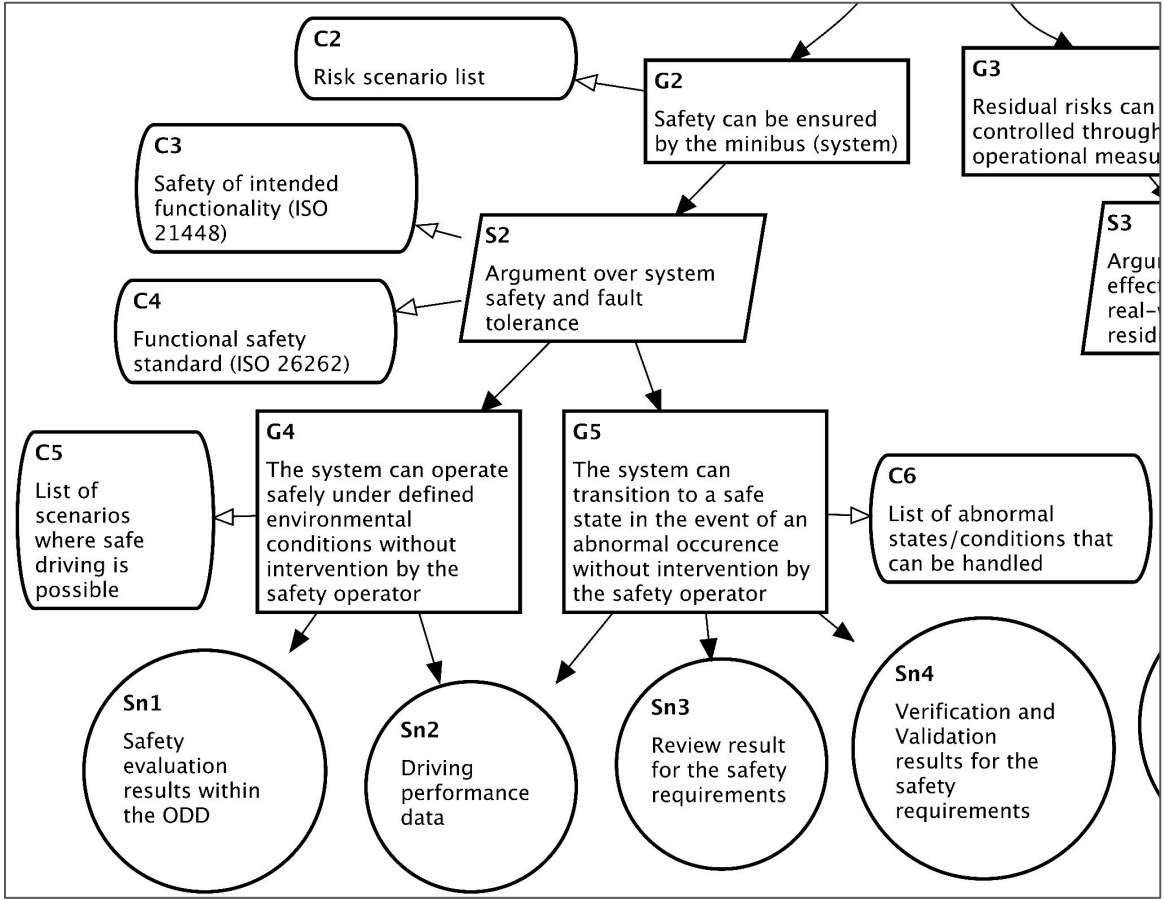
System Safety (G2)

Context nodes:

- C2: Risk scenario list
- C3: SOTIF
- C4: ISO 26262

Sub-goals:

- G4: Safe operation under defined environment conditions
- G5: Transition to safe state during abnormal situations



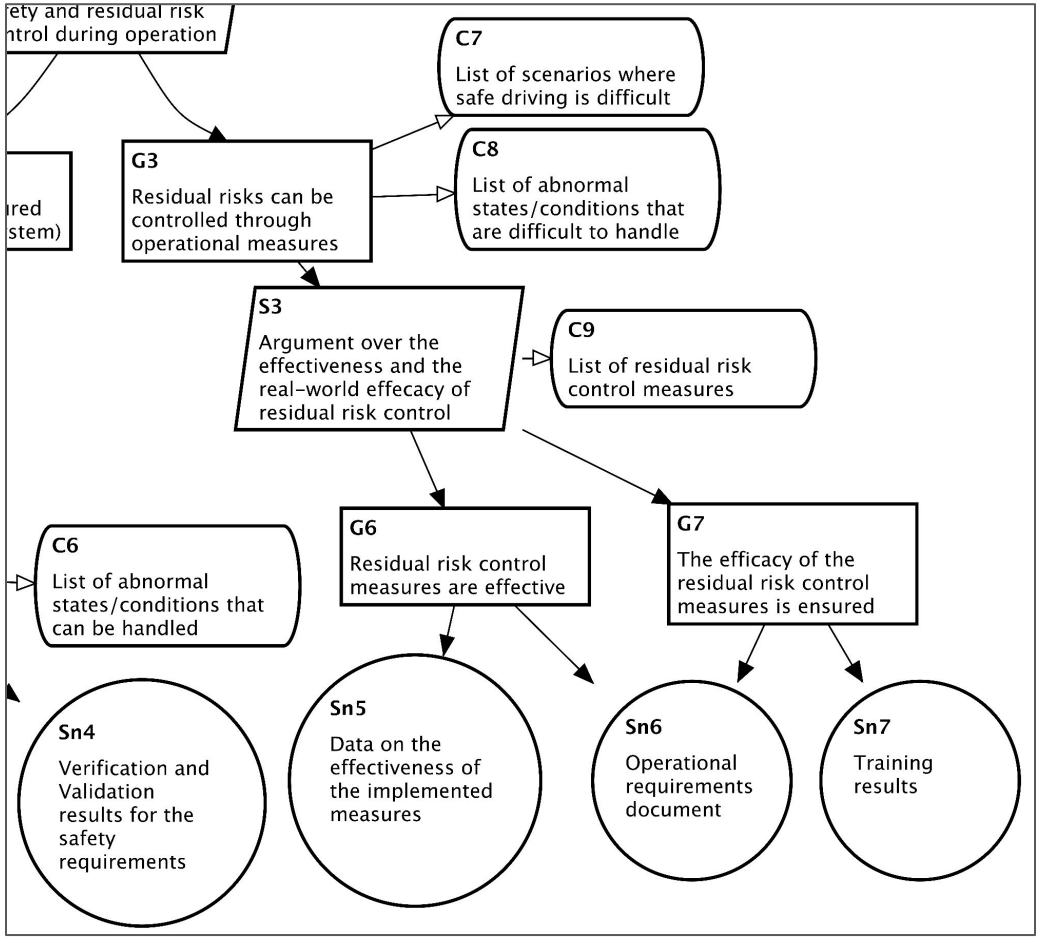
Operational Risk Management (G3)

Sub-goals:

- G6: Effectiveness of risk mitigation measures
- G7: Validation of effectiveness

Contexts:

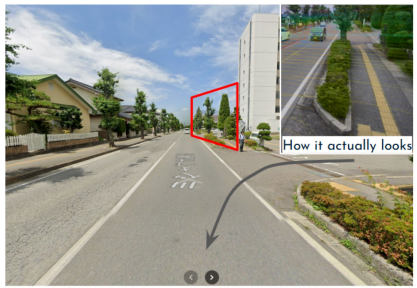
- C7: Challenging driving conditions
- C8: Abnormal state list
- C9: Operational risk control measures



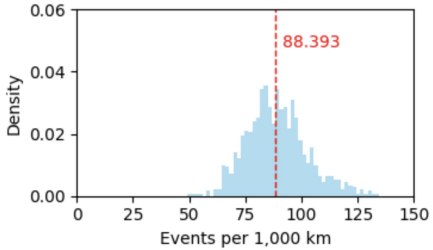
Safety Status Report based on GSN

Objectives

- Identify reasonably foreseeable risks
- Communicate current safety status to stakeholders
- ALARP safety planning



Planting creates the blind spots during turning left



High Risk Area & Hazardous Event Distribution

Contents

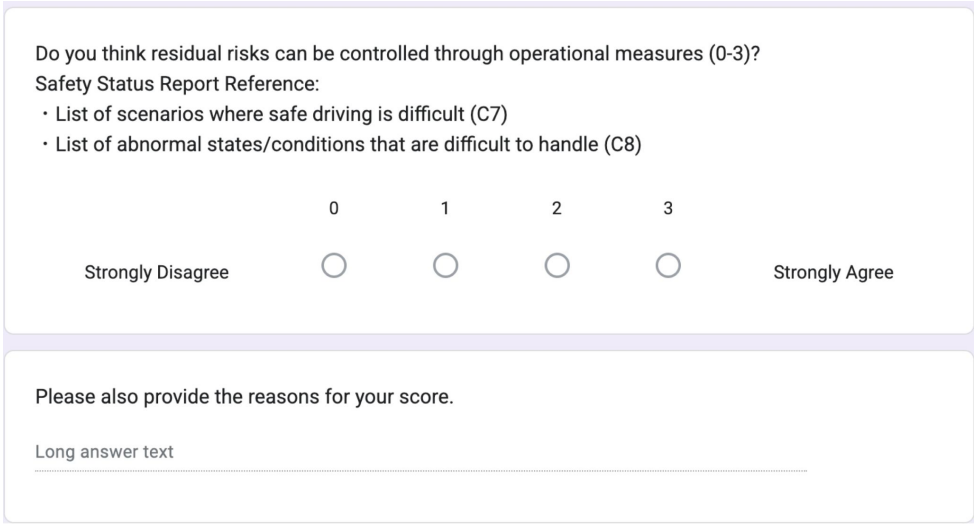
- System Design constraints
- Route restrictions & emergency protocols
- Safety operator procedures
- On-site safety monitors
- Road traffic law compliance

Outcomes

- Path to future full autonomy

Questionnaire Overview

- Jan 13-20, 2025
- 28 TIER IV internal stakeholders directly involved in L4 pilot
- Responses: 21
  1 CxO, 2 Technical Fellows,
  1 Architect, 5 Business Division,
  10 Product Division,
  2 R&D Division
- Rating Scale: 4 point Likert (0-3)
- Questions for G1-G7, S1-S3
  - With open-ended comments



Questionnaire format for Goal G3

Average stakeholder ratings (0-3)

Organization	G1	G2	G3	G4	G5	G6	G7	S1	S2	S3
CxO (n = 1)	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	2.0
Fellow (n = 2)	0.5	2.0	1.0	0.0	0.5	0.5	1.5	1.0	1.5	1.0
Architect (n = 1)	2.0	0.0	2.0	1.0	0.0	2.0	1.0	3.0	3.0	3.0
Business Div. (n = 5)	2.4	1.4	2.4	0.8	1.8	2.0	1.4	3.0	3.0	2.4
Product Div. (n = 10)	1.2	0.9	1.5	1.1	0.9	1.5	1.3	2.8	2.8	2.4
R&D Div. (n = 2)	2.0	1.0	1.5	0.5	1.0	2.0	2.0	2.5	3.0	3.0
Overall (n = 21)	1.6	1.1	1.7	0.9	1.0	1.6	1.4	2.6	2.7	2.3

Role-specific comments

CxO

- Supports overall feasibility

Technical Fellows

- Strongly questions technical safety and evidence sufficiency

Architect

- Generally endorses operational risk mitigation

Business Division

- Positive on early demonstration

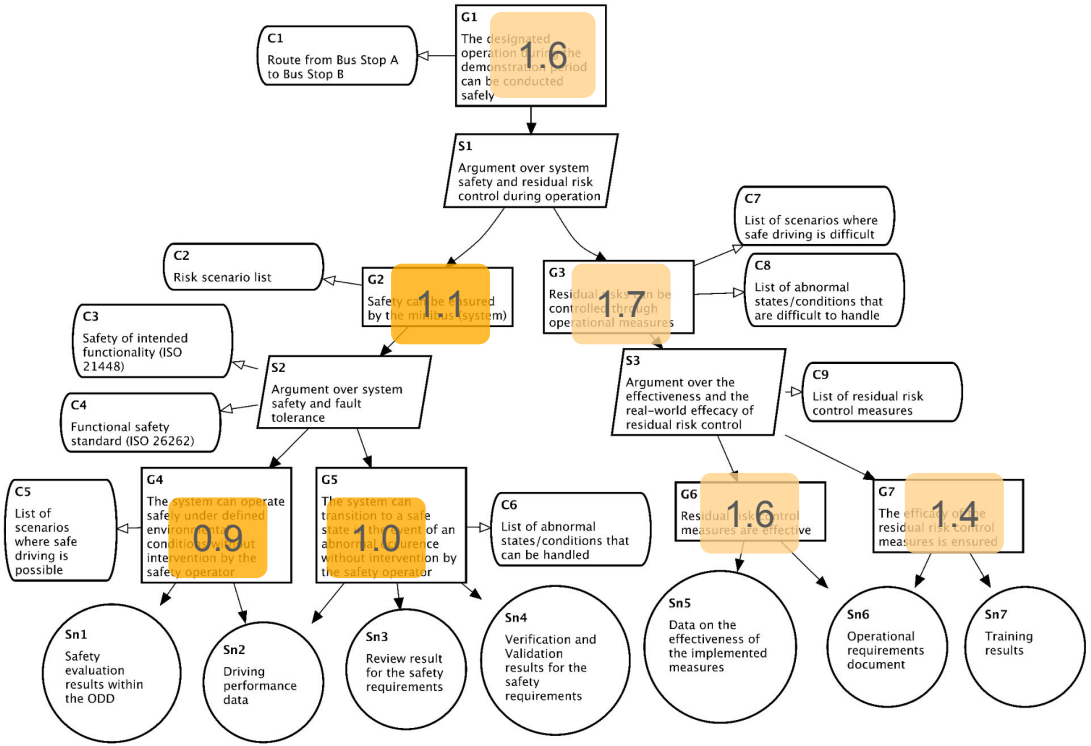
Product Division

- Points out reliance on human intervention

R&D Division

- Calls for deeper analysis of operation-based mitigations and unknown risks

Quantifying Consensus



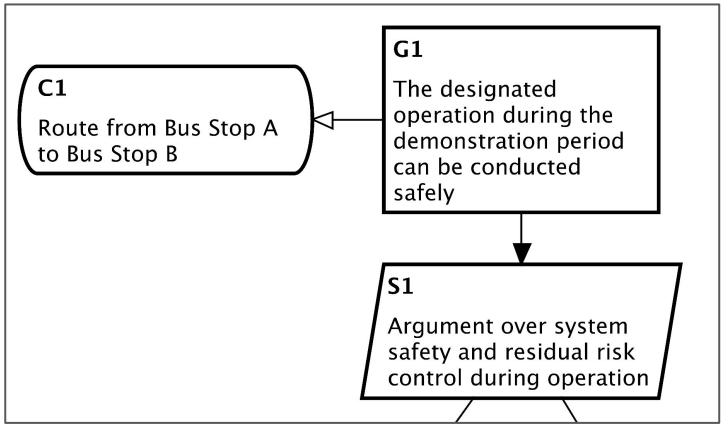
Average scores for Goals ([0,3])

In most cases, each goal's score is higher than its sub goals

- Top-level goal benefit from holistic assessment that naturally extends beyond documented elements
- Specific sub-goals face more rigorous scrutiny of their technical evidence and test coverage

Top-Down and Bottom-Up Views in Assurance Cases

Top-down,
holistic view



Bottom-up,
detailed view

- We can't document everything
- To harmonize these views, we propose Consens Score

Definition (Consensus Score)

Case 1: Leaf Node

For goal G with $AverageRating(G) \in [0,3]$:

$$ConsensusScore(G) = AverageRating(G)/3$$

Normalizes rating to [0,1]

Case 2: Decomposed Goal

When G is decomposed by strategy S:

$$ConsensusScore(G) = (A + B \times C)/2$$

Where:

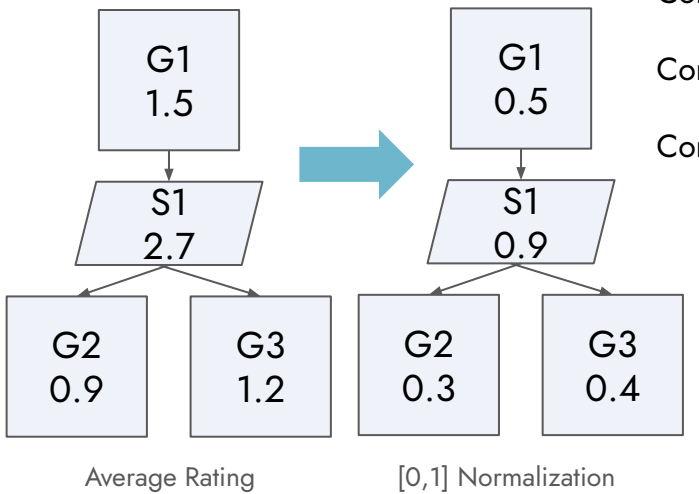
A = AverageRating(G)/3 (normalized goal rating)

B = AverageRating(S)/3 (normalized strategy rating)

C = mean of the consensus scores of G's subgoals

- Recursive definition allows propagation through GSN
- Combines direct evaluation (A) with sub-goal evaluation (B × C)
- All scores normalized to [0,1]

Example



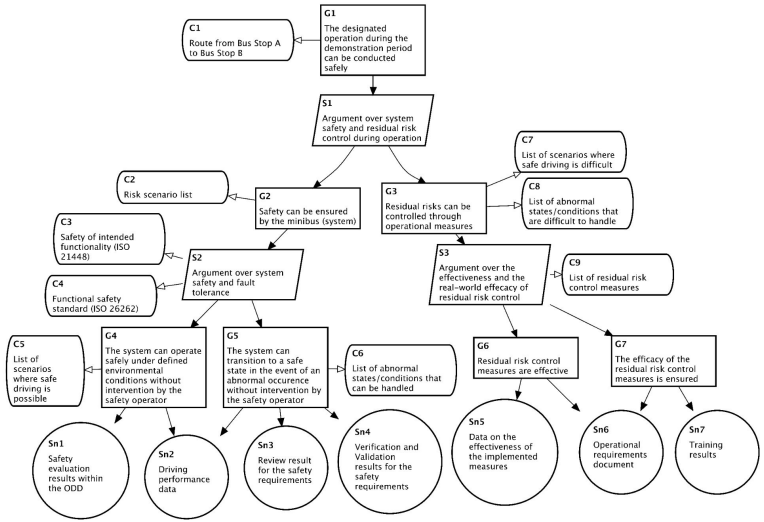
ConsensusScore(G2)=0.3

ConsensusScore(G3)=0.4

ConsensusScore(G1)
=(0.5+0.9×0.35)/2
=0.4075

Consensus Score for TIER IV L4 Demonstration

Node	G1	G2	G3	G4	G5	G6	G7
Consensus Score	0.44	0.33	0.48	0.30	0.33	0.53	0.47



- Agreed with a restricted L4 demonstration
- Remained cautious about deploying full-scale L4 automated driving

Comparison of Consensus Score & CAMs

	Consensus Scoring	CAMs
Focus	Stakeholder agreement & acceptance	Argument validity & technical confidence
Approach	Top-down & bottom-up integration	Bottom-up evidence evaluation
Input	Survey-based stakeholder ratings	Expert judgment & evidence properties
Output	Degree of consensus (0-1 score)	Confidence level & defeater identification
Strength	Identifies acceptance & dissent patterns	Rigorous uncertainty/validity analysis
Best for	Multi-stakeholder decision making	Technical safety verification

- Consensus Score and CAMs are complementary
- In L4 pilot,
  - Not confident in defining detailed parameters to apply CAMs
  - Time and cost constraints

Concluding Remarks

- Stakeholder-oriented framework with Consensus Score
  - Successfully applied in SAE L4 field demonstration
- Key Lessons
  - Transparent Communication
  - Consensus Process Drives Safety
  - Inclusive Stakeholder Engagement
- Future Work
  - Extend to external stakeholders
  - Elaborate Consensus Score
  - Consensus Building based on Confidence Assessment