Safety Argument Framework for Vehicle Autonomy

John Birch
Chief Engineer
Functional Safety

May 17
Agenda

- Safety challenges with autonomy
- Value of an explicit safety argument
- MISRA safety argument model
- Safety argument framework
- Concluding remarks
Safety Challenges with Autonomy
Safety Challenges with Autonomy (SAE Level 3+)

- Safety of the Intended Functionality (SOTIF)
  - Hazardous behaviour not only caused by malfunction
  - Not always clear how system should behave in order to be ‘safe’
  - May be required to trade off one form of hazardous behaviour for another
  - Safety challenge is not just technical but also philosophical and ethical

- No clear definition of acceptable risk
  - Even with ongoing exercise to develop the SOTIF PAS (ISO/PAS 21448) in line with ISO 26262 edition 2

- Required technology at odds with existing standards
  - ‘Non-deterministic’ software
Safety Challenges with Autonomy (SAE Level 3+)

- Safety of the Intended Functionality (SOTIF)
  - Hazardous behaviour not only caused by malfunction.
  - Not always clear how system should behave in order to be 'safe'.
  - May be required to trade off one form of hazardous behaviour for another.
  - Safety challenge is not just technical, but also philosophical and ethical.

- No clear definition of acceptable risk.
  - Even with ongoing exercise to develop the SOTIF PAS (ISO/PAS 21448) in line with ISO 26262 edition 2.

- Required technology at odds with existing standards.
  - ‘Non-deterministic’ software.

Whatever the safety argument is, it needs to be written down!
Value of an Explicit Safety Argument
Value of an Explicit Safety Argument

Adaption of figure from:
Value of an Explicit Safety Argument

Claims

*The autonomous vehicle is acceptably safe for use on public roads*

Argument

Evidence
Value of an Explicit Safety Argument

Claims

The autonomous vehicle is acceptably safe for use on public roads

Argument

Evidence

Test result showing three million miles of incident-free autonomous driving

Successful audit against the requirements of standard x
Value of an Explicit Safety Argument

Claims

The autonomous vehicle is acceptably safe for use on public roads

Argument

???

Evidence

Test result showing three million miles of incident-free autonomous driving

Successful audit against the requirements of standard x
Value of an Explicit Safety Argument

Claims

The autonomous vehicle is acceptably safe for use on public roads.

Evidence

- Successful audit against the requirements of standard x
- Test result showing three million miles of incident-free autonomous driving

Evidence without argument is unexplained
MISRA Safety Case Guidelines: Argument Model
MISRA Safety Case Guidelines

- MISRA (Motor Industry Software Reliability Association) producing a set of guidelines on automotive safety case development
  - Due for publication late 2017
  - Initial scope aligned with ISO 26262 Edition 1
  - Collaborative activity:
MISRA Safety Case Guidelines
Argument Model

- **Rationale**
- **Satisfaction**
- **Means**
- **Environment**

**Product Argument**

**Confidence Argument**
Assurance Argument Framework
Assurance Argument Framework
Item Definition – Autonomous Driver

- Navigation System
- Sensors
- Vehicle surroundings
- Route Request

Autonomous Driver

- Demand
- Propulsion System
- Braking System
- Steering System
Assurance Argument Framework
Item Definition – Autonomous Driver

1. Vehicle surroundings
   - Calculate safety risk at time step \( n \)

2. Determine actuator options for step \( n + 1 \) within physical limitations

3. Predict safety risk for each option

4. Any options with risk lower than the Acceptable Risk Threshold (ART)?
   - Yes (Y)
     - Enact option that would yield optimum route following
   - No (N)
     - Enact option that would yield minimum safety risk

5. Increment time step
Assurance Argument Framework

Functional Safety

The absence of unreasonable risk associated with behaviour of the Autonomous Driver (AD) has been achieved.

‘Unreasonable Risk’ Definition

‘Risk judged to be unacceptable in a certain context according to valid societal moral concepts’

Intended Functions, Malfunctions & Malicious Intent

Argument split according to functionality that is intended, unintended and due to malicious intent.

Intended Behaviour

The absence of unreasonable risk associated with the intended behaviour of the AD has been achieved.

Malfunctioning Behaviour

The absence of unreasonable risk associated with malfunctioning behaviour of the AD has been achieved.

Malicious Intent

The absence of unreasonable risk associated with malicious attack of the AD has been achieved.

See MISRA Safety Case Guidelines…
Assurance Argument Framework

Functional Safety

The absence of unreasonable risk associated with behaviour of the Autonomous Driver (AD) has been achieved

‘Unreasonable Risk’ Definition

‘Risk judged to be unacceptable in a certain context according to valid societal moral concepts’

Intended Functions, Malfunctions & Malicious Intent

Argument split according to functionality that is intended, unintended and due to malicious intent

Intended Behaviour

The absence of unreasonable risk associated with the intended behaviour of the AD has been achieved

Malfunctioning Behaviour

The absence of unreasonable risk associated with malfunctioning behaviour of the AD has been achieved

Malicious Intent

The absence of unreasonable risk associated with malicious attack of the AD has been achieved
Assurance Argument Framework
Functional Safety – Intended Behaviour

Intended Behaviour
The absence of unreasonable risk associated with the intended behaviour of the AD has been achieved

Requirements (REQs)
Argument structured by the specified requirements (REQs)

Requirements Rationale
Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD

Requirements Satisfaction
The AD behaves according to the REQs
Assurance Argument Framework
Functional Safety – Intended Behaviour

Intended Behaviour
The absence of unreasonable risk associated with the intended behaviour of the AD has been achieved

Requirements (REQs)
Argument structured by the specified requirements (REQs)

Requirements Rationale
Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD has been achieved

Requirements Satisfaction
The AD behaves according to the REQs

REQs
REQ 1, REQ 2 etc.

REQs Rationale
Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD

REQs Satisfaction
The AD behaves according to the REQs
Assurance Argument Framework
Functional Safety – Intended Behaviour Rationale

 REQs Rationale
Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD

 Nature of the ART
The ART is specified such that if it is respected then the resulting risk will not be unreasonable

 Respecting the ART when Achievable
Meeting the REQs will ensure that the vehicle behaviour will only fail to respect the ART when prohibited by physical limitations

 Justification for Inability to Respect the ART
Failing to respect the ART due to physical limitations does not yield unreasonable risk

 Societal Acceptance of Limitations
Society accepts that an autonomous vehicle has similar physical limitations to those of a conventional vehicle

 Results of public survey of safety expectations of autonomous vehicles
Assurance Argument Framework
Functional Safety – Intended Behaviour Rationale

REQs Rationale
Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD

Risk Matrix
Acceptable Risk Threshold (ART)
Argument structured by the ART

Nature of the ART
The ART is specified such that if it is respected then the resulting risk will not be unreasonable

Respecting the ART when Achievable
Meeting the REQs will ensure that the vehicle behaviour will only fail to respect the ART when prohibited by physical limitations

Justification for Inability to Respect the ART
Failing to respect the ART due to physical limitations does not yield unreasonable risk

Societal Acceptance of Limitations
Society accepts that an autonomous vehicle has similar physical limitations to those of a conventional vehicle

Results of public survey of safety expectations of autonomous vehicles

Direct Inspection of REQs

Acceptable Risk Threshold
ART Specification

© HORIBA MIRA Ltd. 2017
May 17
Assurance Argument Framework
Functional Safety – Intended Behaviour Rationale

Nature of the ART
The ART is specified such that if it is respected then the resulting risk will not be unreasonable.

Relationship to ‘Model’ Human Driver
The ART is not higher than that which would be exceeded by a ‘model’ human driver.

Justification for ‘Model’ Human Driver
The behaviour of a ‘model’ human driver does not yield a level of risk that would be unreasonable for an Autonomous Driver.

Comparison of ART specification with ‘model’ human driver specification

Results of public survey of safety expectations of autonomous vehicles

‘Model’ Human Driver
‘Model’ Human Driver Behavioural Specification
Assurance Argument Framework
Functional Safety – Intended Behaviour

**Intended Behaviour**

The absence of unreasonable risk associated with the intended behaviour of the AD has been achieved

**Requirements (REQs)**

Argument structured by the specified requirements (REQs)

**Requirements Rationale**

Meeting the REQs yields the absence of unreasonable risk associated with the intended behaviour of the AD

**Requirements Satisfaction**

The AD behaves according to the REQs
Assurance Argument Framework
Functional Safety – Intended Behaviour Satisfaction

**REQs**
REQ 1, REQ 2 etc.

**Test Scenarios**
Virtual and Physical test scenarios

**Acceptable Risk Threshold (ART)**
Argument structured according to the ART

**REQs Satisfaction**
The AD behaves according to the REQs

**Virtual Testing**
During virtual vehicle testing the AD has been shown to only exceed the ART due to physical limitations

**Physical Testing**
During real-world physical vehicle testing the AD has been shown to only exceed the ART due to physical limitations

**Testing Diversity and Number**
The diversity and number of scenarios in which the AD has been tested yields sufficient confidence in having met the REQs

**Risk Matrix**
Risk Matrix Specification

**Acceptable Risk Threshold**
ART Specification

**Test Statistics**
Completed testing statistics

**Industry Measure**
Recognised industry statistical measure for required successful test completion

**Results of virtual vehicle testing**

**Results of physical vehicle testing**

**Comparison of completed testing with industry measure**

**Industry Measure**
Recognised industry statistical measure for required successful test completion
Concluding Remarks
Concluding Remarks

- Safety for autonomy is multi-faceted and challenging
- Important to be able to show structured, explicit reasoning for achievement of safety, particularly to justify residual risk
- Argument may need to be pitched at a higher level of abstraction than would be the case for a ‘conventional system’
- Dynamic safety cases may be required, but automation should not preclude thought!
- Argument likely to require philosophical and ethical reasoning as well as technical
- The devil is in the detail
- Complex problem – not claiming to have the final answer!
Contact details

John Birch
MEng CEng MIMechE
Chief Engineer, Functional Safety

Direct T: +44 (0)2476 355 415
Mobile: +44 (0)7834 158049
Email: John.Birch@horiba-mira.com

HORIBA MIRA Ltd
Watling Street,
Nuneaton, Warwickshire,
CV10 0TU, UK

T: +44 (0)24 7635 5000
F: +44 (0)24 7635 8000

www.horiba-mira.com