Hi-Drive – 1st Summer School, Porto Heli, Greece

Evaluating Causal Effects of SOTIF Triggering Conditions

Christian Neurohr
German Aerospace Center (DLR) e.V.
Institute of Systems Engineering for Future Mobility

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

What is the SOTIF?

- For driving automation relying on environment perception for situational awareness, the intended functionality can cause hazardous behavior despite being free of functional safety faults [ISO 21448]

What is the SOTIF?

- For driving automation relying on environment perception for situational awareness, the intended functionality can cause hazardous behavior despite being free of functional safety faults [ISO 21448]
- The *safety of the intended functionality* (SOTIF) is defined as the absence of unreasonable risk due to hazardous behaviors related to functional insufficiencies [ISO 21448]

What is the SOTIF?

- For driving automation relying on environment perception for situational awareness, the intended functionality can cause hazardous behavior despite being free of functional safety faults [ISO 21448]
- The *safety of the intended functionality* (SOTIF) is defined as the absence of unreasonable risk due to hazardous behaviors related to functional insufficiencies [ISO 21448]
- Functional safety and the SOTIF are complementary aspects of safety

What is the SOTIF?

- For driving automation relying on environment perception for situational awareness, the intended functionality can cause hazardous behavior despite being free of functional safety faults [ISO 21448]
- The *safety of the intended functionality* (SOTIF) is defined as the absence of unreasonable risk due to hazardous behaviors related to functional insufficiencies [ISO 21448]
- Functional safety and the SOTIF are complementary aspects of safety
- The ISO 21448 has been extended to address all levels of driving automation, including automated driving systems (ADSs) at SAE Level ≥ 3

Identification and Evaluation of Triggering Conditions

The ISO 21448 requires the identification and evaluation of potential triggering conditions (Clause 7).

Identification and Evaluation of Triggering Conditions

The ISO 21448 requires the identification and evaluation of potential triggering conditions (Clause 7).

Definition (Triggering Condition, cf. Definition 3.30, ISO 21448)

Specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behavior or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse.

Identification and Evaluation of Triggering Conditions

The ISO 21448 requires the identification and evaluation of potential triggering conditions (Clause 7).

Definition (Triggering Condition, cf. Definition 3.30, ISO 21448)
Specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behavior or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse.

Examples of potential triggering conditions include

- Weather: cloudy, rain, fog, . . .

Identification and Evaluation of Triggering Conditions

The ISO 21448 requires the identification and evaluation of potential triggering conditions (Clause 7).

<u>Definition</u> (Triggering Condition, cf. Definition 3.30, ISO 21448)
Specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behavior or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse.

Examples of potential triggering conditions include

- Weather: cloudy, rain, fog, . . .
- Lighting: glare, night, twilight, . . .

Identification and Evaluation of Triggering Conditions

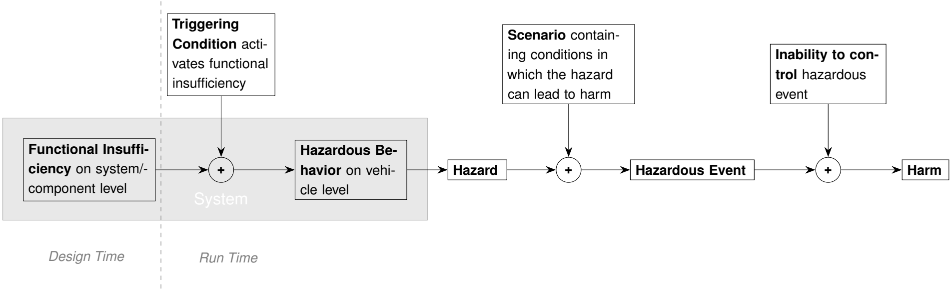
The ISO 21448 requires the identification and evaluation of potential triggering conditions (Clause 7).

Definition (Triggering Condition, cf. Definition 3.30, ISO 21448)
Specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behavior or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse.

Examples of potential triggering conditions include

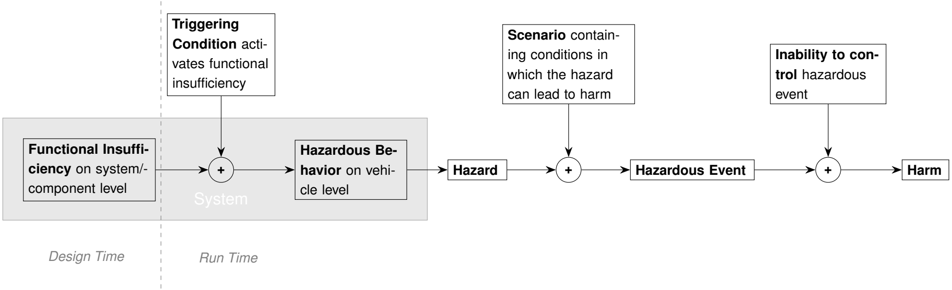
- Weather: cloudy, rain, fog, . . .
- Lighting: glare, night, twilight, . . .
- Road surfaces: asphalt, gravel, potholes, . . .

The ISO 21448's Cause-and-Effect Model



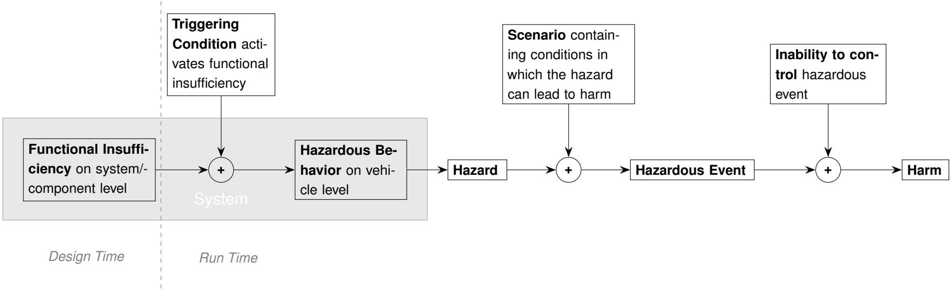
The ISO 21448's Cause-and-Effect Model

- Functional insufficiencies are present in the ADS at design time



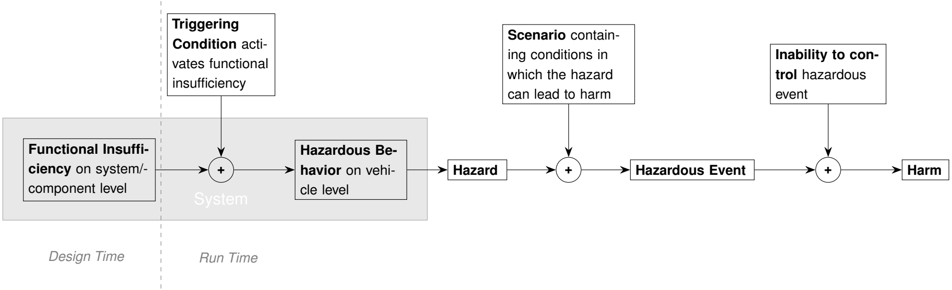
The ISO 21448's Cause-and-Effect Model

- Functional insufficiencies are present in the ADS at design time
- Triggering conditionds activate them during run time



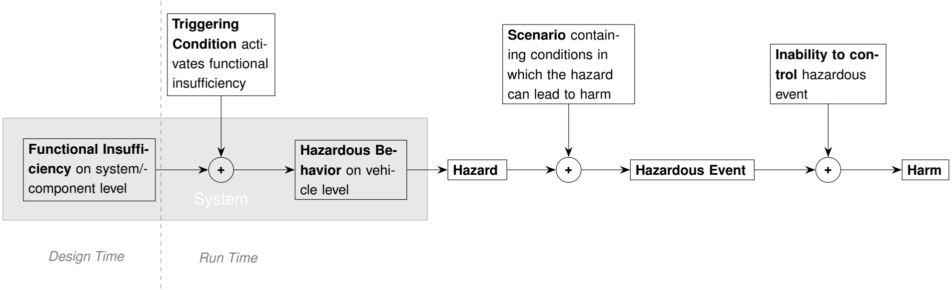
The ISO 21448's Cause-and-Effect Model

- Functional insufficiencies are present in the ADS at design time
- Triggering conditionds activate them during run time
- Due to this, the system exhibits hazardous behavior



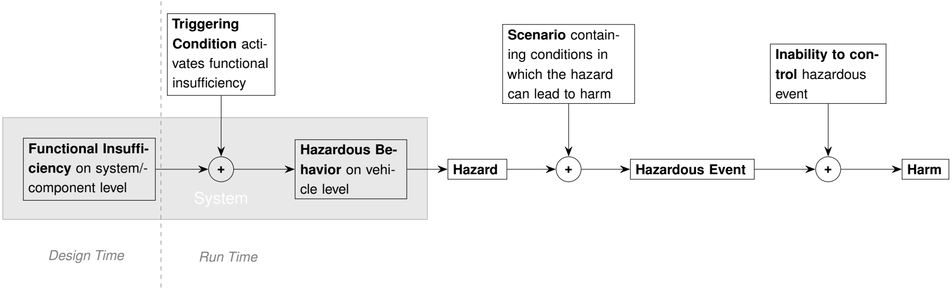
The ISO 21448's Cause-and-Effect Model

- Functional insufficiencies are present in the ADS at design time
- Triggering conditionds activate them during run time
- Due to this, the system exhibits hazardous behavior
- In the correct conditions, the resulting hazard can lead to a hazardous event



The ISO 21448's Cause-and-Effect Model

- Functional insufficiencies are present in the ADS at design time
- Triggering conditionds activate them during run time
- Due to this, the system exhibits hazardous behavior
- In the correct conditions, the resulting hazard can lead to a hazardous event
- If not controlled, this hazardous event can cause harm.



Discussion

- The ISO 21448's cause-and-effect model discretizes the chain from triggering conditions to harm in five stages

Discussion

- The ISO 21448's cause-and-effect model discretizes the chain from triggering conditions to harm in five stages
- These causal chains are linear sequences of events over time (with some additional constraints), possibly leading to harm

Discussion

- The ISO 21448's cause-and-effect model discretizes the chain from triggering conditions to harm in five stages
- These causal chains are linear sequences of events over time (with some additional constraints), possibly leading to harm
- For their qualitative, expert-based evaluation safety analysis techniques can be applied

Discussion

- The ISO 21448’s cause-and-effect model discretizes the chain from triggering conditions to harm in five stages
- These causal chains are linear sequences of events over time (with some additional constraints), possibly leading to harm
- For their qualitative, expert-based evaluation safety analysis techniques can be applied
- For a potential quantitative evaluation, e.g. using the framework of probability theory, problems arise, cf. [Pu22]

Discussion

- The ISO 21448's cause-and-effect model discretizes the chain from triggering conditions to harm in five stages
- These causal chains are linear sequences of events over time (with some additional constraints), possibly leading to harm
- For their qualitative, expert-based evaluation safety analysis techniques can be applied
- For a potential quantitative evaluation, e.g. using the framework of probability theory, problems arise, cf. [Pu22]

Can we use more formal causality frameworks that facilitate the quantitative evaluation of triggering conditions?

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

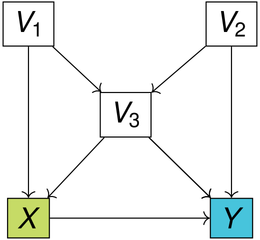
(7) Evaluation of Causal Effects (in `pyAgrum`)

Introduction to Causal Theory

- The systematic investigation of causal questions requires a formal expression of causal relationships

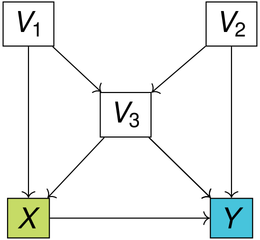
Introduction to Causal Theory

- The systematic investigation of causal questions requires a formal expression of causal relationships
- Causal Inference according to J. Pearl combines graphs and statistics to introduce a formal notion of causality



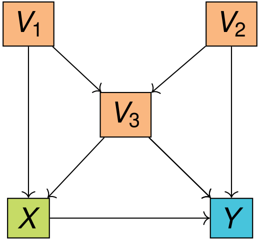
Introduction to Causal Theory

- The systematic investigation of causal questions requires a formal expression of causal relationships
- Causal Inference according to J. Pearl combines graphs and statistics to introduce a formal notion of causality
- The joint probability distribution is directly defined by the graph structure, i.e.



$$P(x, v_1, v_2, v_3, y) = P(v_1) \cdot P(v_2) \cdot P(v_3 | v_1, v_2) \cdot P(x | v_1, v_3) \cdot P(y | v_2, v_3, x)$$

Introduction to Causal Theory



- The systematic investigation of causal questions requires a formal expression of causal relationships
- Causal Inference according to J. Pearl combines graphs and statistics to introduce a formal notion of causality
- The joint probability distribution is directly defined by the graph structure, i.e.

$$P(x, v_1, v_2, v_3, y) = P(v_1) \cdot P(v_2) \cdot P(v_3 | v_1, v_2) \cdot P(x | v_1, v_3) \cdot P(y | v_2, v_3, x)$$

- Explicitly stating assumptions on causal links between variables as a directed, acyclic graphs enables algorithmic confounder analysis

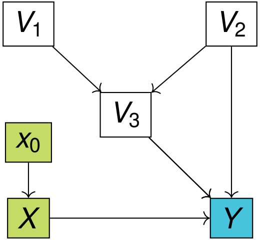
Introduction to Causal Theory (II)

- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

Introduction to Causal Theory (II)

- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$
\begin{aligned}
&P(v_1, v_2, v_3, y \mid \mathrm{do}(X = x_0)) \\
=&P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

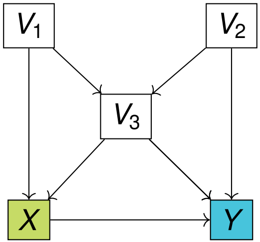


Introduction to Causal Theory (II)

- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$P(v_1, v_2, v_3, y \mid \mathrm{do}(X = x_0))$$
$$= P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)$$

- Using this framework, it is possible to infer causal effects even from observational, non-experimental data



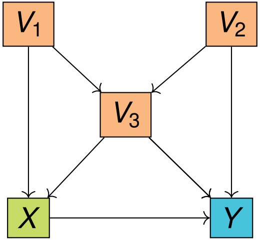
Introduction to Causal Theory (II)

- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$
\begin{aligned}
&P(v_1, v_2, v_3, y \mid \mathrm{do}(X = x_0)) \\
=&P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

- Using this framework, it is possible to infer causal effects even from observational, non-experimental data

$$
\begin{aligned}
&P(Y = y \mid \mathrm{do}(X = x_0)) \\
=&\sum_{v_1, v_2, v_3} P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$



Introduction to Causal Theory (II)

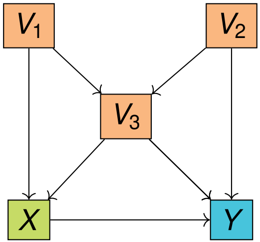
- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$
\begin{aligned}
& P(v_1, v_2, v_3, y \mid \mathrm{do}(X = x_0)) \\
= & P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

- Using this framework, it is possible to infer causal effects even from observational, non-experimental data

$$
\begin{aligned}
& P(Y = y \mid \mathrm{do}(X = x_0)) \\
= & \sum_{v_1, v_2, v_3} P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

- Admissible adjustment sets: $\{V_1, V_3\}, \{V_2, V_3\}, \{V_1, V_2, V_3\}$



Introduction to Causal Theory (II)

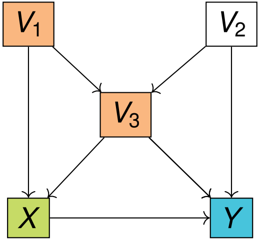
- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$
\begin{aligned}
&P(v_1, v_2, v_3, y \mid \mathrm{do}(X = x_0)) \\
=&P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

- Using this framework, it is possible to infer causal effects even from observational, non-experimental data

$$
\begin{aligned}
&P(Y = y \mid \mathrm{do}(X = x_0)) \\
=&\sum_{v_1, v_2, v_3} P(v_1) \cdot P(v_2) \cdot P(v_3 \mid v_1, v_2) \cdot P(y \mid v_2, v_3, x_0)
\end{aligned}
$$

- Admissible adjustment sets: $\{V_1, V_3\}, \{V_2, V_3\}, \{V_1, V_2, V_3\}$



Introduction to Causal Theory (II)

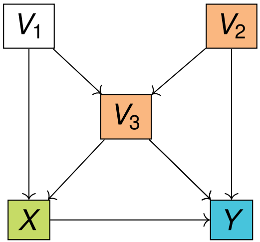
- Interventions (known from randomized controlled trials) are expressed using so-called do-operator

$$\begin{aligned}&P(v_1, v_2, v_3, y \,|\, \mathrm{do}(X = x_0))\\ =&P(v_1) \cdot P(v_2) \cdot P(v_3|v_1, v_2) \cdot P(y|v_2, v_3, x_0)\end{aligned}$$

- Using this framework, it is possible to infer causal effects even from observational, non-experimental data

$$\begin{aligned}&P(Y = y \,|\, \mathrm{do}(X = x_0))\\ =&\sum_{v_1, v_2, v_3} P(v_1) \cdot P(v_2) \cdot P(v_3|v_1, v_2) \cdot P(y|v_2, v_3, x_0)\end{aligned}$$

- Admissible adjustment sets:
  $\{V_1, V_3\}, \{V_2, V_3\}, \{V_1, V_2, V_3\}$



Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

Criticality Metrics for Automated Driving

Definition (Definition 1, Ne21)

Criticality (of a traffic situation) is the combined risk of the involved actors when the traffic situation is continued.

Criticality Metrics for Automated Driving

Definition (Definition 1, Ne21)

Criticality (of a traffic situation) is the combined risk of the involved actors when the traffic situation is continued.

Definition (Section 1, We23)

A (scene-level) criticality metric is a function $\kappa : \mathcal{S} \to \mathcal{O} \subseteq \mathbb{R} \cup \{\pm\infty\}$ that measures, for a given traffic scene $S \in \mathcal{S}$, aspects of criticality.

Criticality Metrics for Automated Driving

Definition (Definition 1, Ne21)

Criticality (of a traffic situation) is the combined risk of the involved actors when the traffic situation is continued.

Definition (Section 1, We23)

A (scene-level) criticality metric is a function $\kappa : \mathcal{S} \to \mathcal{O} \subseteq \mathbb{R} \cup \{\pm\infty\}$ that measures, for a given traffic scene $S \in \mathcal{S}$, aspects of criticality.

- Most criticality metrics only seek to quantify certain aspects of criticality, e.g. *Time*, *Space*, *Dynamics*, *Perception*, *Environment*, ...

Criticality Metrics for Automated Driving

Definition (Definition 1, Ne21)

Criticality (of a traffic situation) is the combined risk of the involved actors when the traffic situation is continued.

Definition (Section 1, We23)

A (scene-level) criticality metric is a function $\kappa : \mathcal{S} \to \mathcal{O} \subseteq \mathbb{R} \cup \{\pm\infty\}$ that measures, for a given traffic scene $S \in \mathcal{S}$, aspects of criticality.

- Most criticality metrics only seek to quantify certain aspects of criticality, e.g. *Time*, *Space*, *Dynamics*, *Perception*, *Environment*, ...
- Scenario-level criticality metrics extend this definition from scenes to scenarios, i.e. taking into account the actual temporal evolution

Criticality Metrics for Automated Driving

Definition (Definition 1, Ne21)

Criticality (of a traffic situation) is the combined risk of the involved actors when the traffic situation is continued.

Definition (Section 1, We23)

A (scene-level) criticality metric is a function $\kappa : \mathcal{S} \to \mathcal{O} \subseteq \mathbb{R} \cup \{\pm\infty\}$ that measures, for a given traffic scene $S \in \mathcal{S}$, aspects of criticality.

- Most criticality metrics only seek to quantify certain aspects of criticality, e.g. *Time*, *Space*, *Dynamics*, *Perception*, *Environment*, . . .
- Scenario-level criticality metrics extend this definition from scenes to scenarios, i.e. taking into account the actual temporal evolution
- Examples include Time-To-Collision, Post-Encroachment Time, Required Acceleration

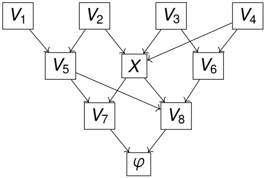
Presentation Structure

Application of Causal Theory to Automotive Safety

- How can the framework provided by causal theory be leveraged?

Application of Causal Theory to Automotive Safety

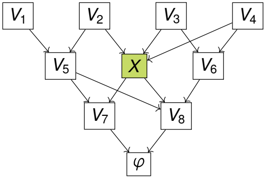
- How can the framework provided by causal theory be leveraged?
- Koopmann et al. define a *causal relation* as a pair consisting of a causal graph together with a *context* (a set of statements within a suitable traffic domain ontology) [Ko22]



Context:
- $\mathtt{a}$ exists and is of type $\mathtt{C}_1$
- $\mathtt{b}$ exists and is of type $\mathtt{C}_1$, but does not have property $\mathtt{p}_1$
- $\mathtt{c}$ exists and is of type $\mathtt{C}_2$
- It has to hold that $\mathtt{a.p}_2 < \mathtt{b.p}_2$.

Application of Causal Theory to Automotive Safety

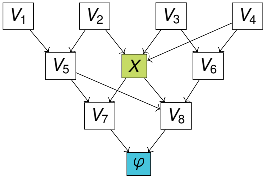
- How can the framework provided by causal theory be leveraged?
- Koopmann et al. define a *causal relation* as a pair consisting of a causal graph together with a *context* (a set of statements within a suitable traffic domain ontology) [Ko22]
- A triggering condition *tc* is modeled as a binary random variable's value $X \in \{tc, \neg tc\}$



Context:
- $\mathtt{a}$ exists and is of type $\mathtt{C}_1$
- $\mathtt{b}$ exists and is of type $\mathtt{C}_1$, but does not have property $\mathtt{p}_1$
- $\mathtt{c}$ exists and is of type $\mathtt{C}_2$
- It has to hold that $\mathtt{a}.\mathtt{p}_2 < \mathtt{b}.\mathtt{p}_2$.

Application of Causal Theory to Automotive Safety

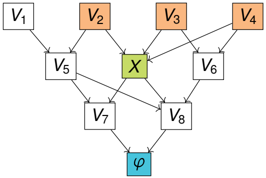
- How can the framework provided by causal theory be leveraged?
- Koopmann et al. define a *causal relation* as a pair consisting of a causal graph together with a *context* (a set of statements within a suitable traffic domain ontology) [Ko22]
- A triggering condition *tc* is modeled as a binary random variable's value $X \in \{tc, \neg tc\}$
- As sink of the causal graph, a criticality metric φ is used for measurement



Context:
- $\mathtt{a}$ exists and is of type $\mathtt{C}_1$
- $\mathtt{b}$ exists and is of type $\mathtt{C}_1$, but does not have property $\mathtt{p}_1$
- $\mathtt{c}$ exists and is of type $\mathtt{C}_2$
- It has to hold that $\mathtt{a.p}_2 < \mathtt{b.p}_2$.

Application of Causal Theory to Automotive Safety

- How can the framework provided by causal theory be leveraged?
- Koopmann et al. define a *causal relation* as a pair consisting of a causal graph together with a *context* (a set of statements within a suitable traffic domain ontology) [Ko22]
- A triggering condition tc is modeled as a binary random variable's value $X \in \{tc, \neg tc\}$
- As sink of the causal graph, a criticality metric φ is used for measurement



Context:
- $\mathtt{a}$ exists and is of type $\mathtt{C}_1$
- $\mathtt{b}$ exists and is of type $\mathtt{C}_1$, but does not have property $\mathtt{p}_1$
- $\mathtt{c}$ exists and is of type $\mathtt{C}_2$
- It has to hold that $\mathtt{a}.\mathtt{p}_2 < \mathtt{b}.\mathtt{p}_2$.

Modeling of Causal Relations

- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*

Modeling of Causal Relations

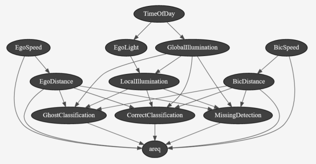
- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*
- these triggering conditions are each modeled as a value of a discrete random variable corresponding to a node in the graph

Modeling of Causal Relations

- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*
- these triggering conditions are each modeled as a value of a discrete random variable corresponding to a node in the graph
- Required longitudinal acceleration $a_{long,req}$ is used as criticality metric

Modeling of Causal Relations

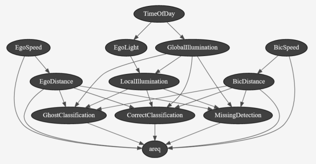
- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*
- these triggering conditions are each modeled as a value of a discrete random variable corresponding to a node in the graph
- Required longitudinal acceleration $a_{long,req}$ is used as criticality metric



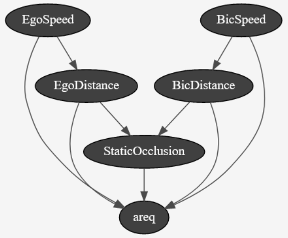
(a) Low Local Illumination

Modeling of Causal Relations

- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*
- these triggering conditions are each modeled as a value of a discrete random variable corresponding to a node in the graph
- Required longitudinal acceleration $a_{long,req}$ is used as criticality metric



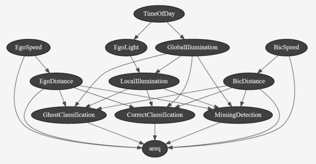
(a) Low Local Illumination



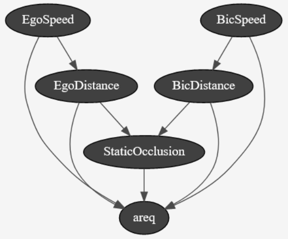
(b) Static Occlusion

Modeling of Causal Relations

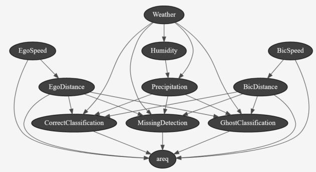
- As examples, we consider three potential triggering conditions: *low local illumination*, *static occlusion*, and *heavy rain*
- these triggering conditions are each modeled as a value of a discrete random variable corresponding to a node in the graph
- Required longitudinal acceleration $a_{long,req}$ is used as criticality metric



(a) Low Local Illumination

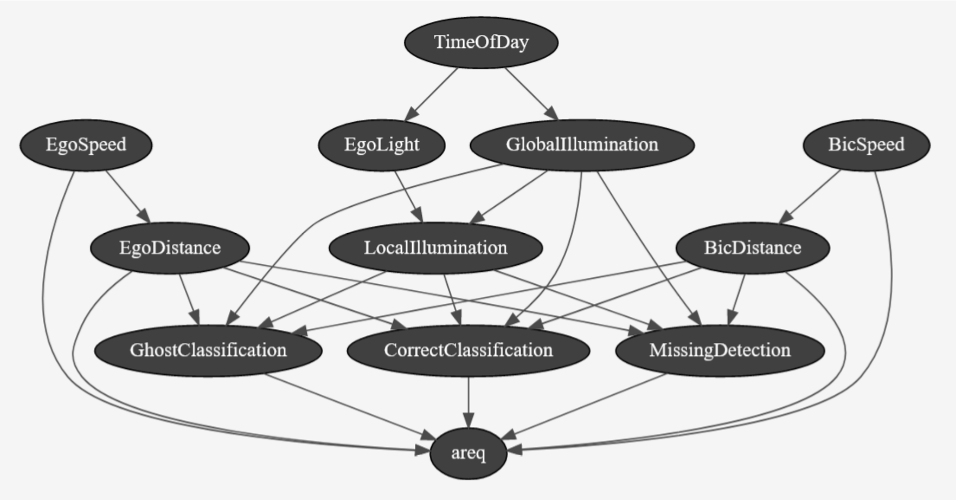


(b) Static Occlusion



(c) Heavy Rain

Modeling of Causal Relation: Local Illumination



Example Context: Urban Intersection Scenario with Occlusion

We rely on the Automotive Urban Traffic Ontology (A.U.T.O.) [We22], in particular the 6-Layer Model [Sc21], to structure the context.

Example Context: Urban Intersection Scenario with Occlusion

We rely on the Automotive Urban Traffic Ontology (A.U.T.O.) [We22], in particular the 6-Layer Model [Sc21], to structure the context.

Layer	Property
(L1) Road Network and Traffic Guidance Objects	Road network consists of a 3-armed urban junction.
(L2) Roadside Structures	Roadside structures may exist and are not further constrained.
(L3) Temporary Modifications of (L1) and (L2)	No temporary modifications to layers 1 and 2.
(L4) Dynamic Objects	Ego vehicle (going straight), bicyclist (turning left, ignoring right-of-way), static (potentially occluding) object
(L5) Environmental Conditions	Environmental conditions exist and remain unconstrained.
(L6) Digital Information	No digital information.

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

Derivation of Requirements on Data Collection

The most important case is real-world data acquisition with an ADS-operated vehicle.

Derivation of Requirements on Data Collection

The most important case is real-world data acquisition with an ADS-operated vehicle.

(i) Nodes can be modeled as discrete random variables which are measured during test drives

Derivation of Requirements on Data Collection

The most important case is real-world data acquisition with an ADS-operated vehicle.

(i) Nodes can be modeled as discrete random variables which are measured during test drives

(ii) There is sufficient data to instantiate the causal relation, i.e. to estimate the conditional probability distributions

Derivation of Requirements on Data Collection

The most important case is real-world data acquisition with an ADS-operated vehicle.

(i) Nodes can be modeled as discrete random variables which are measured during test drives

(ii) There is sufficient data to instantiate the causal relation, i.e. to estimate the conditional probability distributions

If only a single causal effect is to be quantified, these requirements can be relaxed (admissible adjustment sets)

Derivation of Requirements on Data Collection

The most important case is real-world data acquisition with an ADS-operated vehicle.

(i) Nodes can be modeled as discrete random variables which are measured during test drives

(ii) There is sufficient data to instantiate the causal relation, i.e. to estimate the conditional probability distributions

If only a single causal effect is to be quantified, these requirements can be relaxed (admissible adjustment sets)

(iii) The context is recognizable during a test drive (to activate data collection)

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

(i’) Nodes can be modeled as discrete random variables which are logged during a simulation run

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

(i') Nodes can be modeled as discrete random variables which are logged during a simulation run

(ii') [Sufficient synthetic data can easily be generated]

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

(i') Nodes can be modeled as discrete random variables which are logged during a simulation run

(ii') [Sufficient synthetic data can easily be generated]

(iii') The context is realizable in the simulation

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

(i') Nodes can be modeled as discrete random variables which are logged during a simulation run

(ii') [Sufficient synthetic data can easily be generated]

(iii') The context is realizable in the simulation

(iv') The simulation environment and models are *valid* in the context

Derivation of Requirements on Data Collection (II)

For synthetic data generation, requirements are slightly different.

(i') Nodes can be modeled as discrete random variables which are logged during a simulation run

(ii') [Sufficient synthetic data can easily be generated]

(iii') The context is realizable in the simulation

(iv') The simulation environment and models are *valid* in the context

Real knowledge **can not** be gained from a simulation.

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

Logical Scenario for Example Context

- As to estimate the necessary conditional probabilities, we generate synthetic data using `CARLA` in a logical scenario

Logical Scenario for Example Context

- As to estimate the necessary conditional probabilities, we generate synthetic data using `CARLA` in a logical scenario

Parameter	**Range**
ego start position (m)	$[-58, -33] \times [-29, -28]$
ego target position (m)	$[50, 55] \times [-29, -28]$
ego target speed (km/h)	$[25, 60]$
bicyclist start position (m)	$[31, 32] \times [3, 15]$
bicyclist target position (m)	$[-50, -45] \times [-34, -33]$
bicyclist target speed (km/h)	$[10, 25]$
Dimension of O (parking cars)	$\{0, 1, 2, 3, 4, 5, 6, 7\}$
Position of O (m)	$[2, 20] \times ([-35, -34] \cup [-26, -25])$
Weather	$\{$Clear, Heavy Rain, $\dots\}$

Logical Scenario for Example Context

- As to estimate the necessary conditional probabilities, we generate synthetic data using `CARLA` in a logical scenario
- For simplicity, we draw 900 parameter combinations uniformly from the parameter ranges

Parameter	**Range**
ego start position (m)	$[-58, -33] \times [-29, -28]$
ego target position (m)	$[50, 55] \times [-29, -28]$
ego target speed (km/h)	$[25, 60]$
bicyclist start position (m)	$[31, 32] \times [3, 15]$
bicyclist target position (m)	$[-50, -45] \times [-34, -33]$
bicyclist target speed (km/h)	$[10, 25]$
Dimension of O (parking cars)	$\{0, 1, 2, 3, 4, 5, 6, 7\}$
Position of O (m)	$[2, 20] \times ([-35, -34] \cup [-26, -25])$
Weather	{Clear, Heavy Rain, . . . }

Logical Scenario for Example Context

- As to estimate the necessary conditional probabilities, we generate synthetic data using `CARLA` in a logical scenario
- For simplicity, we draw 900 parameter combinations uniformly from the parameter ranges
- The *ego* is operated by a simple extension of `CARLA`'s basic agent using a front camera with perception trained using `YOLOv4`

Parameter	**Range**
ego start position (m)	$[-58, -33] \times [-29, -28]$
ego target position (m)	$[50, 55] \times [-29, -28]$
ego target speed (km/h)	$[25, 60]$
bicyclist start position (m)	$[31, 32] \times [3, 15]$
bicyclist target position (m)	$[-50, -45] \times [-34, -33]$
bicyclist target speed (km/h)	$[10, 25]$
Dimension of O (parking cars)	$\{0, 1, 2, 3, 4, 5, 6, 7\}$
Position of O (m)	$[2, 20] \times ([-35, -34] \cup [-26, -25])$
Weather	{Clear, Heavy Rain, ...}

Logical Scenario for Example Context

- As to estimate the necessary conditional probabilities, we generate synthetic data using `CARLA` in a logical scenario
- For simplicity, we draw 900 parameter combinations uniformly from the parameter ranges
- The *ego* is operated by a simple extension of `CARLA`'s basic agent using a front camera with perception trained using `YOLOv4`
- The bicyclist is based on an aggressive basic agent and does not respect the *ego*'s right of way

Parameter	**Range**
ego start position (m)	$[-58, -33] \times [-29, -28]$
ego target position (m)	$[50, 55] \times [-29, -28]$
ego target speed (km/h)	$[25, 60]$
bicyclist start position (m)	$[31, 32] \times [3, 15]$
bicyclist target position (m)	$[-50, -45] \times [-34, -33]$
bicyclist target speed (km/h)	$[10, 25]$
Dimension of O (parking cars)	$\{0, 1, 2, 3, 4, 5, 6, 7\}$
Position of O (m)	$[2, 20] \times ([-35, -34] \cup [-26, -25])$
Weather	$\{$Clear, Heavy Rain, $\ldots\}$

Visualization of Concrete Simulation Runs

Visualization of Concrete Simulation Runs (ii)

Visualization of Concrete Simulation Runs (iii)

Presentation Structure

(1) SOTIF and Triggering Conditions

(2) Causal Theory Framework

(3) Criticality Metrics for Automated Driving

(4) Application of Causal Theory to Automotive Safety

(5) Derivation of Requirements on Data Collection

(6) Synthetic Data Generation (in `CARLA`)

(7) Evaluation of Causal Effects (in `pyAgrum`)

Quantification of Causal Effects

Definition (ACE & RCE, cf. Definition 4,Ko22)
For a causal relation that is sufficiently instantiated in its context, the **average** respectively **relative causal effect** of a binary random variable $X = \{tc, \neg tc\}$ **on** a criticality metric φ can be defined as

$$ACE(X, \varphi) := E(\varphi \mid \mathrm{do}(X = tc)) - E(\varphi \mid \mathrm{do}(X = \neg tc)) \,,$$

$$RCE(X, \varphi) := \frac{E(\varphi \mid \mathrm{do}(X = tc))}{E(\varphi \mid \mathrm{do}(X = \neg tc))} \,.$$

Quantification of Causal Effects

Definition (ACE & RCE, cf. Definition 4,Ko22)
For a causal relation that is sufficiently instantiated in its context, the **average** respectively **relative causal effect** of a binary random variable $X = \{tc, \neg tc\}$ **on** a criticality metric φ can be defined as

$$ACE(X,\varphi) := E(\varphi \mid \mathrm{do}(X = tc)) - E(\varphi \mid \mathrm{do}(X = \neg tc)) \,,$$

$$RCE(X,\varphi) := \frac{E(\varphi \mid \mathrm{do}(X = tc))}{E(\varphi \mid \mathrm{do}(X = \neg tc))} \,.$$

Many other quantities representing causal effects are conceivable.

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*

Quantity	Value
ACE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	$0.41 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	1.14
ACE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	$0.44 m/s^2$
RCE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	1.13
ACE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	$0.85 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	1.28
ACE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	$2.01 m/s^2$
RCE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	1.82
ACE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	$-0.03 m/s^2$
RCE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	0.99

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*
- A significant causal effect of local illumination is observed

Quantity	Value
ACE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	$0.41 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	1.14
ACE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	$0.44 m/s^2$
RCE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	1.13
ACE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	$0.85 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	1.28
ACE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	$2.01 m/s^2$
RCE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	1.82
ACE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	$-0.03 m/s^2$
RCE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	0.99

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*
- A significant causal effect of local illumination is observed
- The causal effect of a static occlusion is even stronger

Quantity	Value
ACE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	$0.41 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	1.14
ACE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	$0.44 m/s^2$
RCE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	1.13
ACE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	$0.85 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	1.28
ACE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	$2.01 m/s^2$
RCE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	1.82
ACE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	$-0.03 m/s^2$
RCE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	0.99

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*
- A significant causal effect of local illumination is observed
- The causal effect of a static occlusion is even stronger
- Precipitation has no causal effect on criticality in this context (as it is not implemented in `CARLA` ..)

Quantity	**Value**
$ACE(\text{LocalIllumination: Low} \leftarrow \text{Medium}, a_{long,req})$	$0.41 m/s^2$
$RCE(\text{LocalIllumination: Low} \leftarrow \text{Medium}, a_{long,req})$	1.14
$ACE(\text{LocalIllumination: Medium} \leftarrow \text{High}, a_{long,req})$	$0.44 m/s^2$
$RCE(\text{LocalIllumination: Medium} \leftarrow \text{High}, a_{long,req})$	1.13
$ACE(\text{LocalIllumination: Low} \leftarrow \text{High}, a_{long,req})$	$0.85 m/s^2$
$RCE(\text{LocalIllumination: Low} \leftarrow \text{High}, a_{long,req})$	1.28
$ACE(\text{StaticOcclusion: True} \leftarrow \text{False}, a_{long,req})$	$2.01 m/s^2$
$RCE(\text{StaticOcclusion: True} \leftarrow \text{False}, a_{long,req})$	1.82
$ACE(\text{Precipitation: High} \leftarrow \text{Low}, a_{long,req})$	$-0.03 m/s^2$
$RCE(\text{Precipitation: High} \leftarrow \text{Low}, a_{long,req})$	0.99

Evaluation of Causal Effects in Example Logical Scenario

- A preliminary implementation using `pyAgrum` enables the evaluation of causal effects such as *ACE* and *RCE*
- A significant causal effect of local illumination is observed
- The causal effect of a static occlusion is even stronger
- Precipitation has no causal effect on criticality in this context (as it is not implemented in `CARLA` ..)
- Verdict: the ADS fails in the simulation; the simulation fails regarding precipitation

Quantity	Value
ACE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	$0.41 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ Medium, $a_{long,req}$)	1.14
ACE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	$0.44 m/s^2$
RCE(LocalIllumination: Medium $\leftarrow$ High, $a_{long,req}$)	1.13
ACE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	$0.85 m/s^2$
RCE(LocalIllumination: Low $\leftarrow$ High, $a_{long,req}$)	1.28
ACE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	$2.01 m/s^2$
RCE(StaticOcclusion: True $\leftarrow$ False, $a_{long,req}$)	1.82
ACE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	$-0.03 m/s^2$
RCE(Precipitation: High $\leftarrow$ Low, $a_{long,req}$)	0.99

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality
4. Select context(s) in which the causal graphs are possible

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality
4. Select context(s) in which the causal graphs are possible
5. Derive requirements on data collection

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality
4. Select context(s) in which the causal graphs are possible
5. Derive requirements on data collection
6. Collect data with active ADS during test drives

   Optional: Generate synthetic data with active ADS in simulation

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality
4. Select context(s) in which the causal graphs are possible
5. Derive requirements on data collection
6. Collect data with active ADS during test drives

   Optional: Generate synthetic data with active ADS in simulation
7. Instantiate causal relations with data (estimate probability distributions)

Summary: Evaluation of Triggering Conditions

1. Identify potential triggering conditions (TCs)
2. Choose criticality metrics to measure their effects
3. Build causal graphs how TCs lead to increased criticality
4. Select context(s) in which the causal graphs are possible
5. Derive requirements on data collection
6. Collect data with active ADS during test drives

   Optional: Generate synthetic data with active ADS in simulation
7. Instantiate causal relations with data (estimate probability distributions)
8. Evaluate causal effects of TCs on criticality metrics

Discussion (II)

- Which methods are available for the identification of triggering conditions?

Discussion (II)

- Which methods are available for the identification of triggering conditions?
- Is the ISO 21448's cause-and-effect model already sufficient for the (qualitative/quantitative) evaluation of triggering conditions?

Discussion (II)

- Which methods are available for the identification of triggering conditions?
- Is the ISO 21448's cause-and-effect model already sufficient for the (qualitative/quantitative) evaluation of triggering conditions?
- What problems could arise when trying to evaluate triggering conditions using causal inference?

Discussion (II)

- Which methods are available for the identification of triggering conditions?
- Is the ISO 21448's cause-and-effect model already sufficient for the (qualitative/quantitative) evaluation of triggering conditions?
- What problems could arise when trying to evaluate triggering conditions using causal inference?
- Could criticality metrics be considered surrogate measures for *risk of harm*? If so, which ones?

Discussion (II)

- Which methods are available for the identification of triggering conditions?
- Is the ISO 21448's cause-and-effect model already sufficient for the (qualitative/quantitative) evaluation of triggering conditions?
- What problems could arise when trying to evaluate triggering conditions using causal inference?
- Could criticality metrics be considered surrogate measures for *risk of harm*? If so, which ones?
- Could computer simulations be faithfully used for ADS safeguarding, if their validity is established?

Thank you for the attention.

Contact:

Dr. Christian Neurohr
German Aerospace Center (DLR) e.V.
Institute of Systems Engineering for Future Mobility
christian.neurohr@dlr.de

References

[Pearl] J. Pearl, 'Causal Inference in Statistics: An Overview', in Statistics Surveys, 2009

[Ne21] Neurohr et al., 'Criticality Analysis for the Verification and Validation of Automated Vehicles, in IEEE Access, 2021.

[Sc21] M. Scholtes et al., '6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment', in IEEE Access, 2021

[ISO21448] International Organization for Standardization, 'ISO 21448: Road vehicles – Safety of the intended functionality', 2022.

[We22] Westhofen et al., 'Using Ontologies for the Formalization and Recognition of Criticality for Automated Driving', in IEEE Open Journal of Intelligent Transportation Systems, 2022.

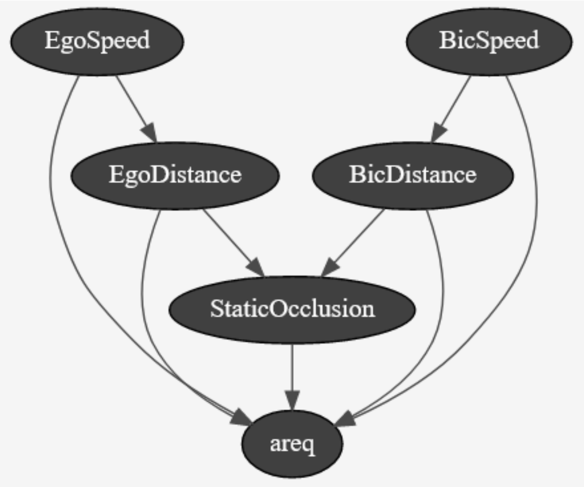
References (II)

[Ko22] Koopmann et al., 'Grasping Causality for the Explanation of Criticality for Automated Driving, arXiv preprint, 2022.

[Pu23] L. Putze et al., 'On Quantification for SOTIF Validation of Automated Driving Systems', 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, USA.

[We23] Westhofen et al., 'Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art'. Archives of Computational Methods in Engineering, 2023.

Modeling of Causal Relation: Static Occlusion



Modeling of Causal Relation: Heavy Rain