Applying Causal Models for the Safety Analysis of Automated Transport Systems

Lina Putze

German Aerospace Center (DLR) e.V.

Institute of Systems Engineering for Future Mobility

Women in Data Science Worldwide | Regensburg

DLR

Automated Transport Systems

- Technological advances pave the way for increasing levels of automation
- These systems operate in an complex operational design domain that is difficult to fully characterize
- With increasing automation the human as redundant monitoring instance is omitted
- Decision making procedures based on AI algorithms are often treated as black-box

Challenges for Automated Systems

- Current **accident data bases** are not representative for automated systems
- Statistical evidence of safety by **distance-based approaches** is <u>not</u> feasible
  - Every update of the system requires a new evaluation
  - Ethically problematic

2022 Germany	Accidents	Overall Distance	Distance between two accidents	Accident-free distance required for evidence (confidence: 95%)
Injuries	289 672	$7.1*10^{11}$ km	$2.6*10^{6}$ km	$7,34*10^{6}$ km
Fatalities	2788	$7.1*10^{11}$ km	$2.5*10^{8}$ km	$7,63*10^{8}$ km

Source: Statistisches Bundesamt (Destatis), 2023

- Established safety processes are primarily concerned with hazardous events caused by component **faults and failures**
- For automated systems relying on situational awareness, the **specified functionality** itself can cause hazardous situations despite the absence functional safety faults

Hazard and Risk Analysis

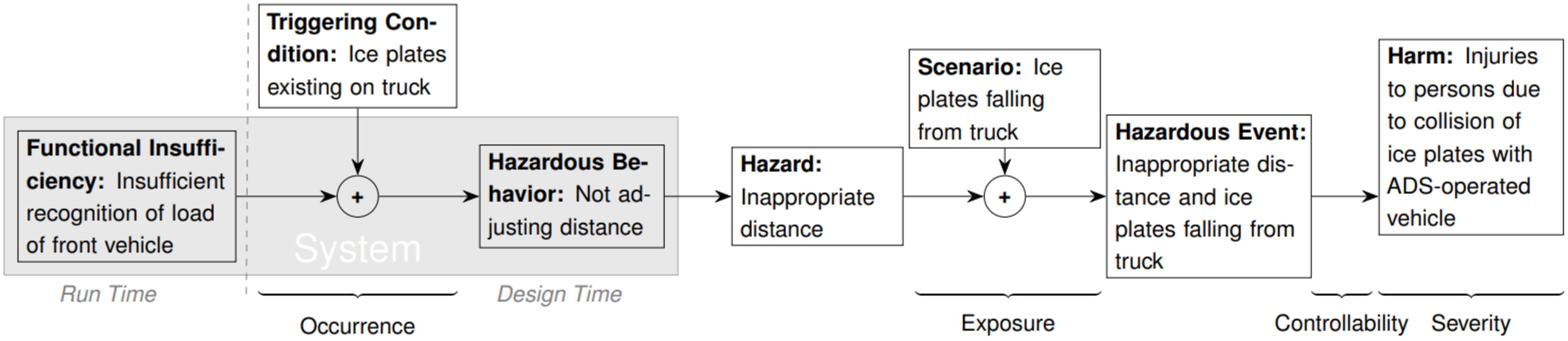
Definition (Risk Analysis, ISO/IEC Guide 51):

Systematic use of available information to identify hazards and to estimated their risks

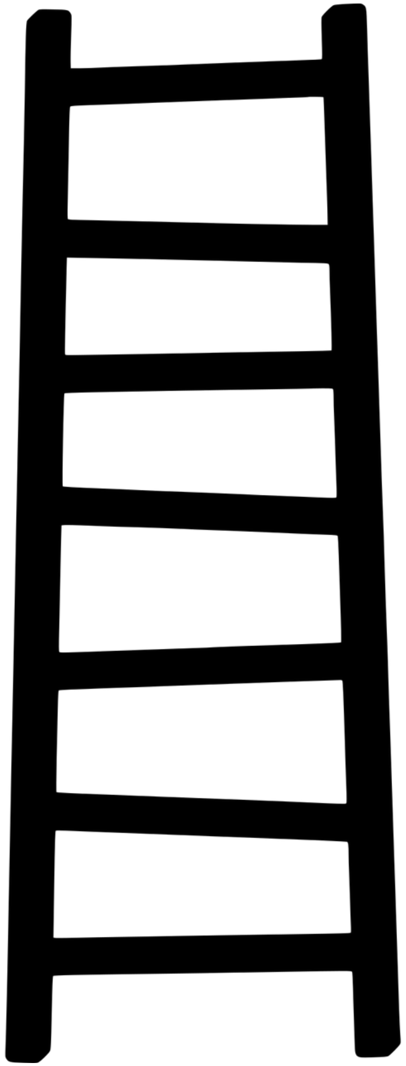
Definition (Hazard, ISO/IEC Guide 51):

Potential source of harm.

➢ A hazard and risk analysis aims at identifying and evaluating potential **causes** of a harm

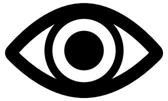


Hierachy of Causality



Counterfactuals: ‘If X had occurred, what would have been Y?’

Intervention: ‘If I do X, how will it change Y’



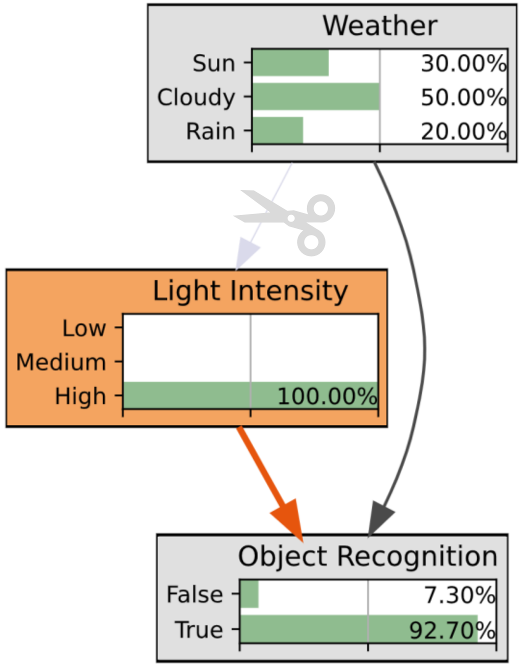
Association: ‘If I see X, what does it tell me about Y?’

Causal Theory

- Causal theory according to J. Pearl provides a **formal notion** of causality by combining graphs with Bayesian statistics
- The **joint probability distribution** can be estimated based on the causal structure:

$$P(o, l, w) = P(o|l, w) \cdot P(l|w) \cdot P(w)$$

- The **do-operator** $do(X = x)$ simulates an intervention by deleting incoming edges defining X and setting X=x for all other variables
- The **do-calculus** provides means to estimate a causal effect based on observational, non-experimental data
- A set of variables is **admissible for adjustment** if it is sufficient to estimate the causal effect



Causal Safety Analysis

- Modelling
- Instantiation
- Verification
- Evaluation of Causal Effects
- Investigation of Safety Measures

Causal Safety Analysis

- Modelling
- Instantiation
- Verification
- Evaluation of Causal Effects
- Investigation of Safety Measures

Modelling the Causal Structure

Given a hazard…

i. define a context
ii. find suitable criticality metrics
iii. model the system
iv. model the environment

The **context** of a causal structure defines a set of constrains on the **existence** and **properties of objects** in suitable ontology.

	Layer	Property
L1	Street Layer	A road shall exist, no further constraints
L2	Traffic Infrastructure	unconstrained
L3	Temporal Modifications	No temporal modifications
L4	Dynamic Objects	An Ego vehicle and another object shall exist
L5	Environment Conditions	unconstrained
L6	Digital Information	unconstrained

Modelling the Causal Structure

Given a hazard…

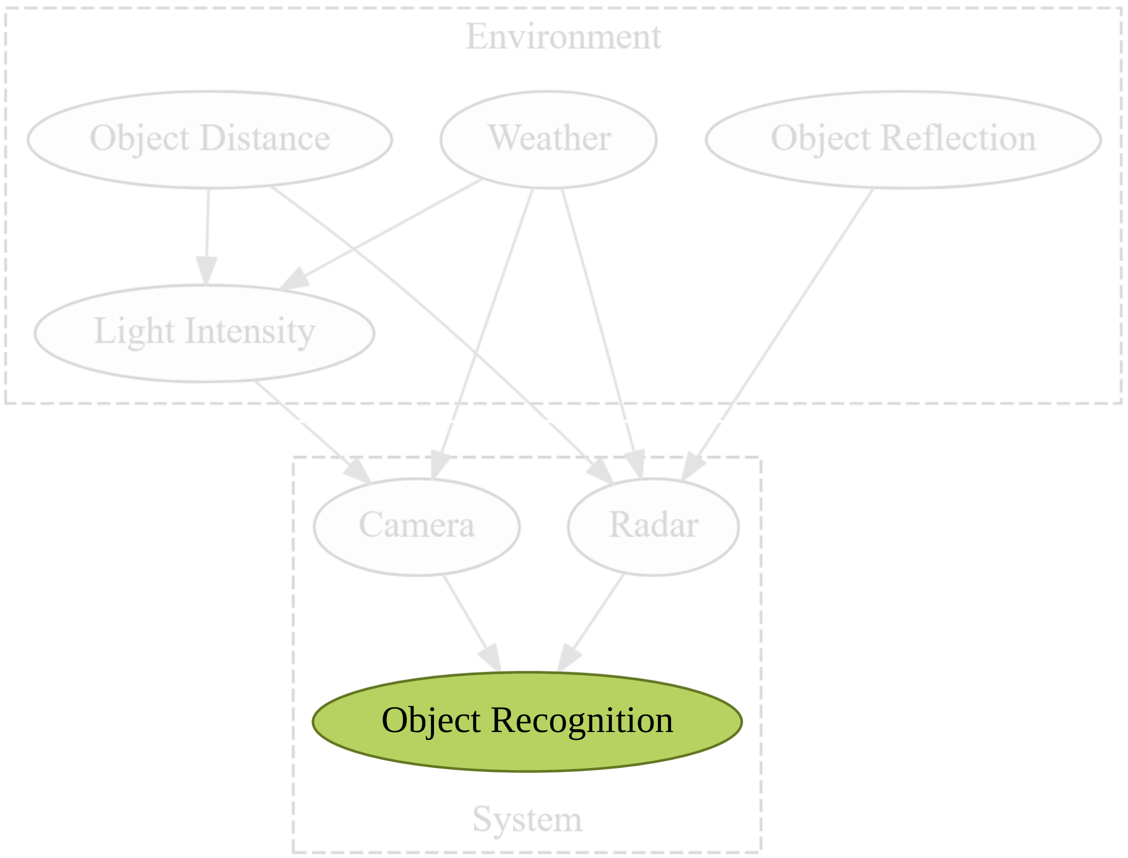
i. define a context

ii. find suitable criticality metrics

iii. model the system

iv. model the environment

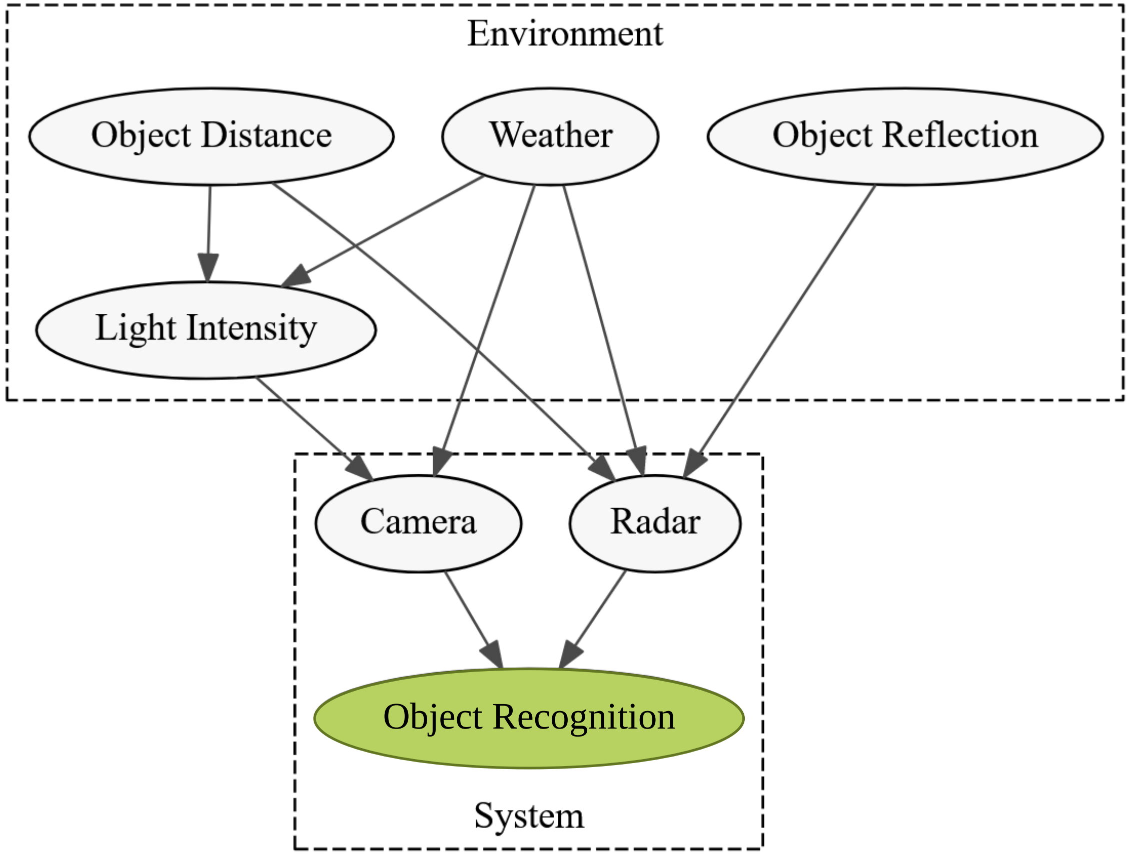
A **criticality metric** is a function that estimates **aspects of criticality** in a scene or scenario.



Modelling the Causal Structure

Given a hazard…

i. define a context
ii. find suitable criticality metrics
iii. model the system
iv. model the environment



Causal Safety Analysis

Modelling

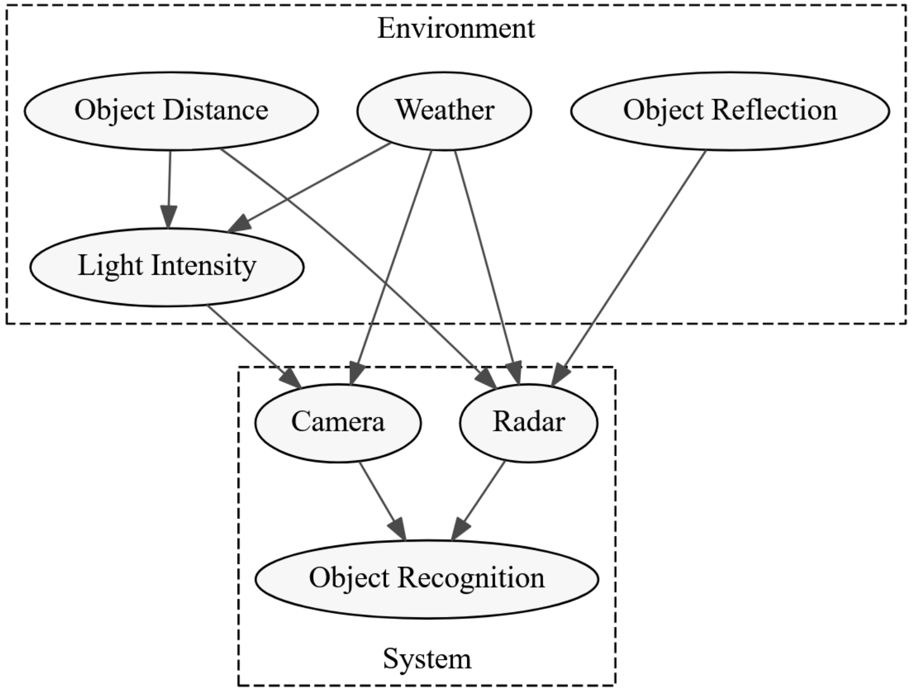
Instantiation

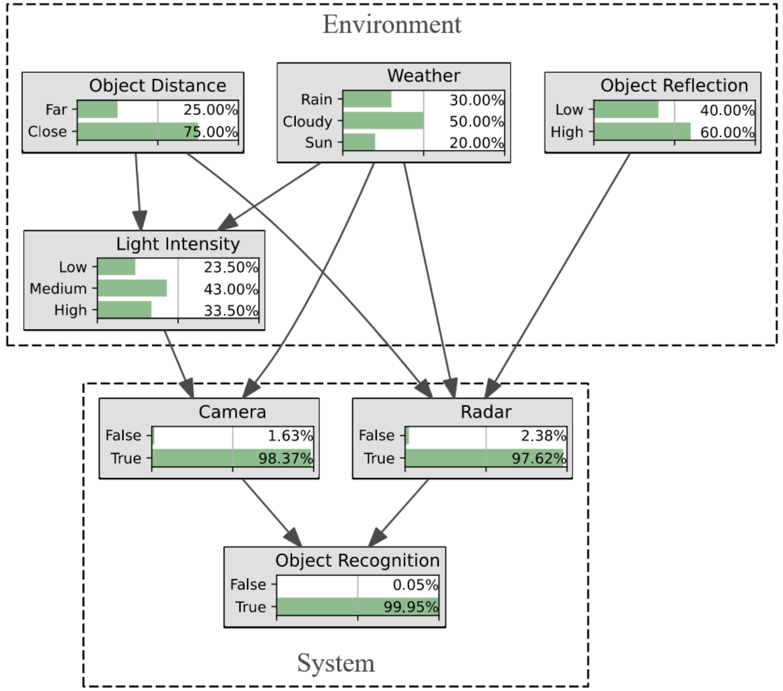
Verification

Evaluation of Causal Effects

Investigation of Safety Measures

Instantiation

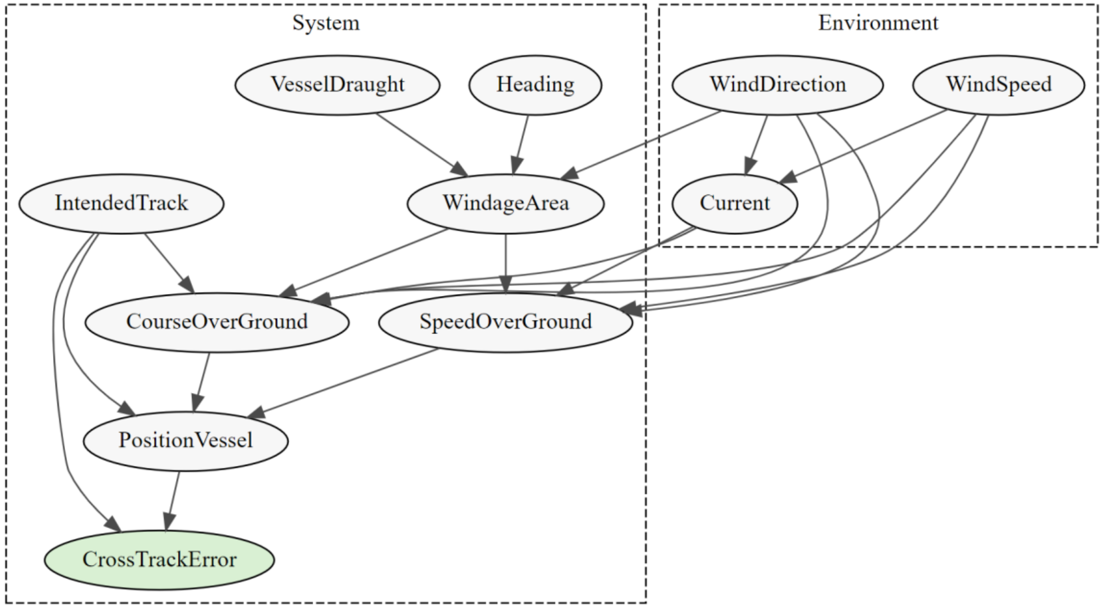




A causal relation is called **partially instantiated** w.r.t. a subset of nodes N by a Dataset D, if the CPDs of the nodes in N are instantiated by D.

It is called **instantiated** for a node X w.r.t. N by D, if it is partially initiated w.r.t. N by D and N contains the criticality metric φ and at least one adjustment for the causal effect of X on φ.

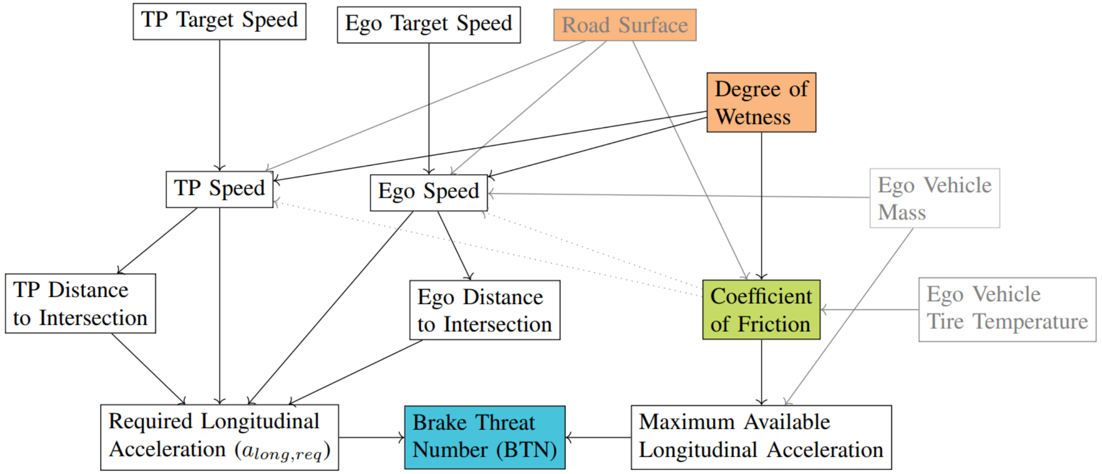
Instantiation - Real World Data

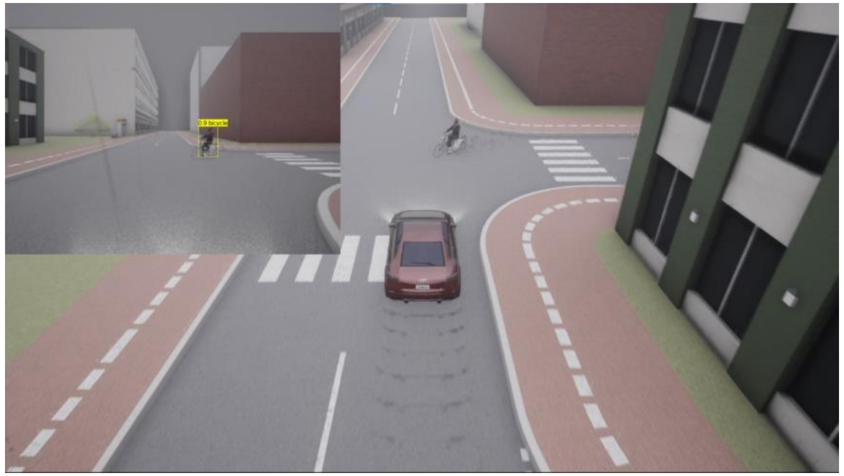


<u>Requirements:</u>

i. Nodes can be described as discrete random variables which are **measured** during test drives or can be obtain from existing data
ii. There is **sufficient data** for the instantiation of the causal relation
iii. The **context is observable** during the test drives and in existing data used

Instantiation - Simulation





<u>Requirements:</u>

i. Nodes can be described as discrete random variables which are **logged during simulation runs**
ii. The context is **realizable** in the simulation
iii. The simulation environment and models are **valid** in the context

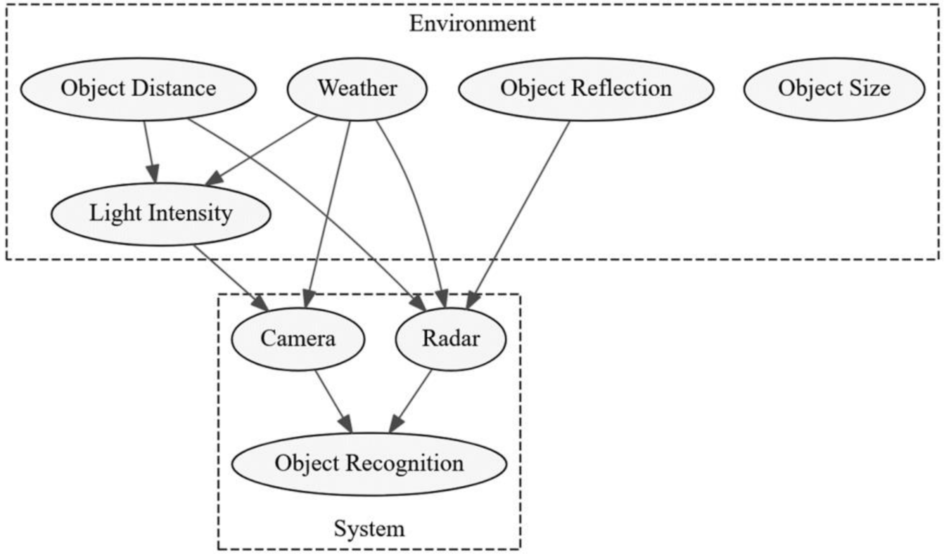
Causal Safety Analysis

- Modelling
- Instantiation
- **Verification**
- Evaluation of Causal Effects
- Investigation of Safety Measures

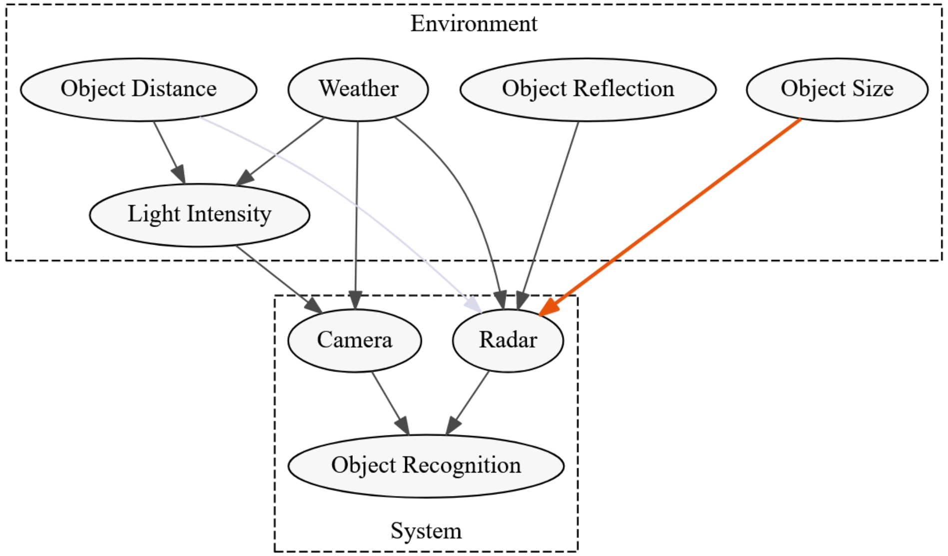
Verification of Causal Assumptions

Goal: Evaluation of the modeling quality

Modelled Causalities:



Real Causalities:



Object Distance	
Far	Close
0.2500	0.7500

Object Size	
Small	Large
0.2500	0.7500

- Kullback-Leibler divergence: $KL(P(M_{Model})|Q(M_{Real})) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right) = 0{,}0134 > 0$
- Hellinger distance: $H^2(P(M_{Model})|Q(M_{Real})) = \frac{1}{2}\sqrt{\sum_x(\sqrt{P(x)} - \sqrt{Q(x)})^2} = 0{,}066 > 0$

Causal Safety Analysis

Modelling

Instantiation

Verification

Evaluation of Causal Effects

Investigation of Safety Measures

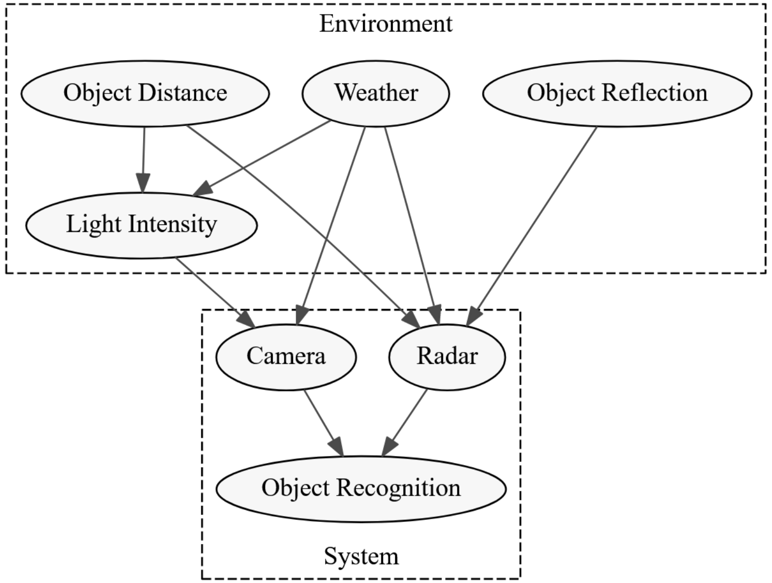
Evaluation of Causal Effects

<u>**Goal:**</u> Investigation of the causal influence …

…of **single causal factors**

…of **combinations** of causal factors

…via specific **paths**



Evaluation of Causal Effects

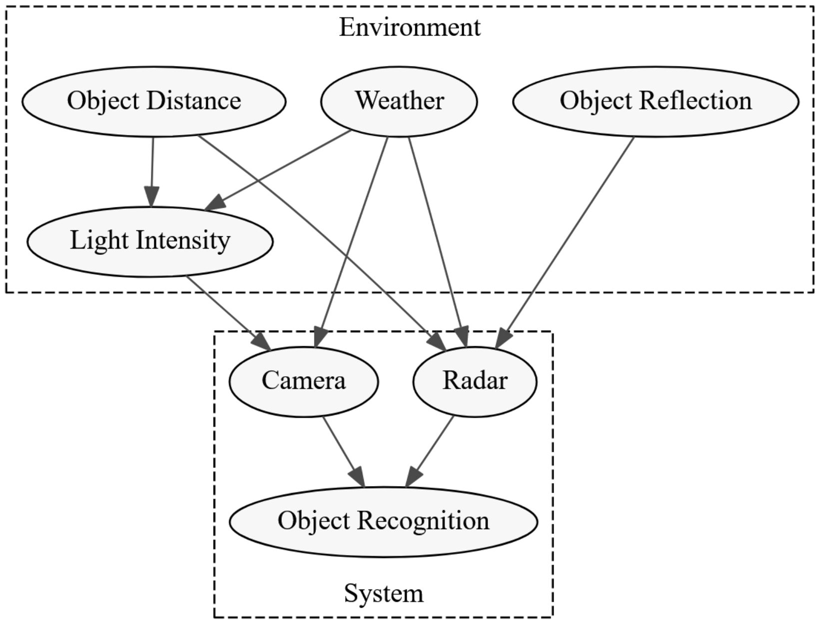
<u>Goal:</u> Investigation of the causal influence of **single causal factors**

- Average Causal Effect:

$$ACE = P(Y|do(X = x)) - P(Y|do(X = x_{ref}))$$

- Relative Causal Effect:

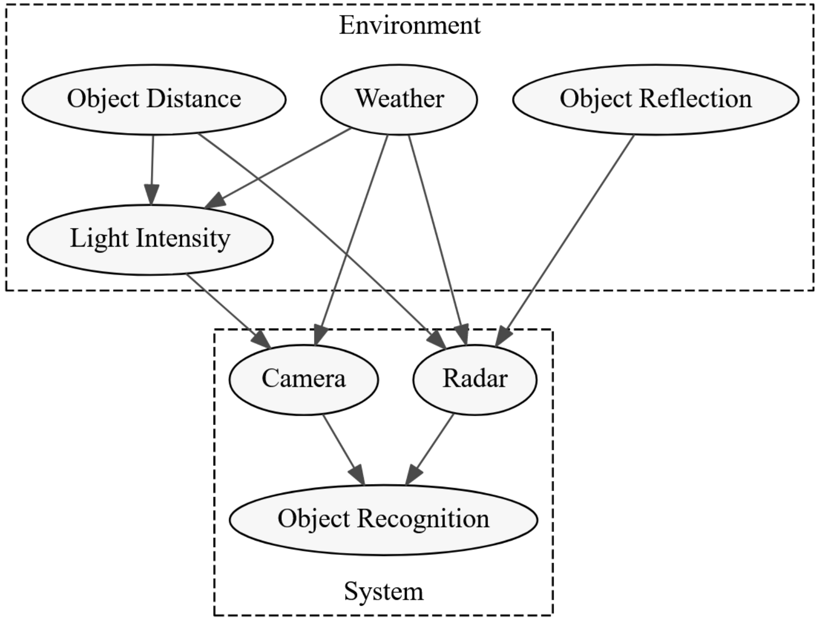
$$\mathrm{R}CE = \frac{P(Y|do(X=x))}{P(Y|do(X=x_{ref}))}$$



Evaluation of Causal Effects

<u>Goal:</u> Investigation of the causal influence of **combinations** of causal factors

- Causal models allow for the investigation of multiple combined interventions
- The causal metrics ACE, RCE and IRRW can be adopted for multiple interventions by replacing the variable being intervened on by a vector of variables



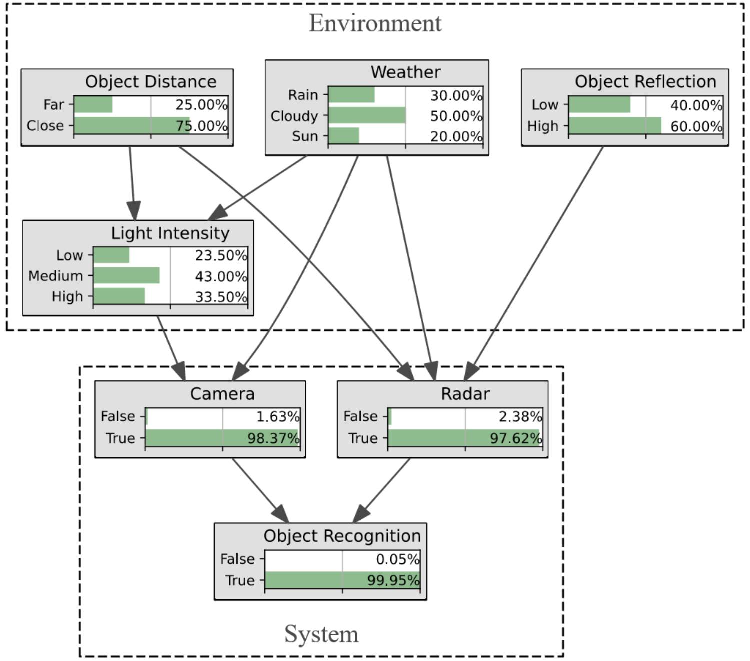
Causal Safety Analysis

Investigation of Safety Measures

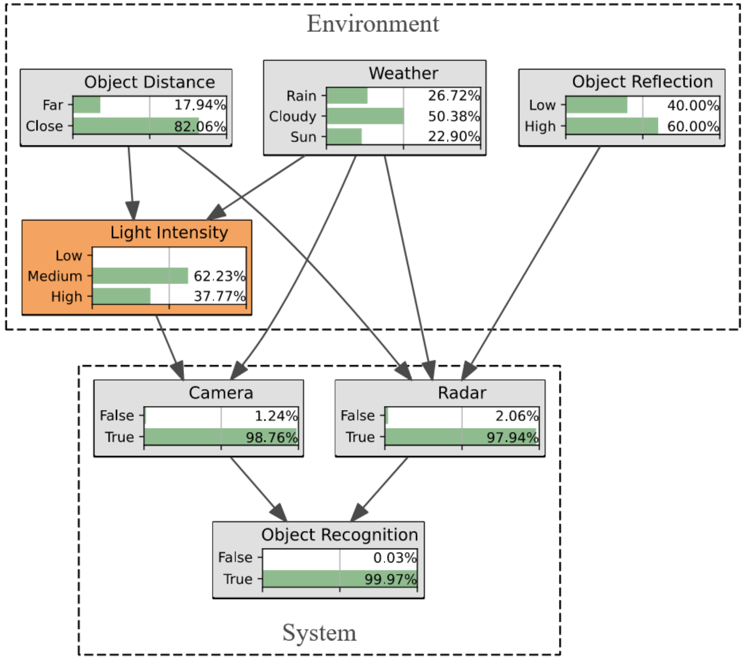
Categories of Safety Measures:

- Technical modifications
- Adjustment of the behavior/ dynamics
- Restriction of the operational design domain
- Adaption of the communication
- Structural changes

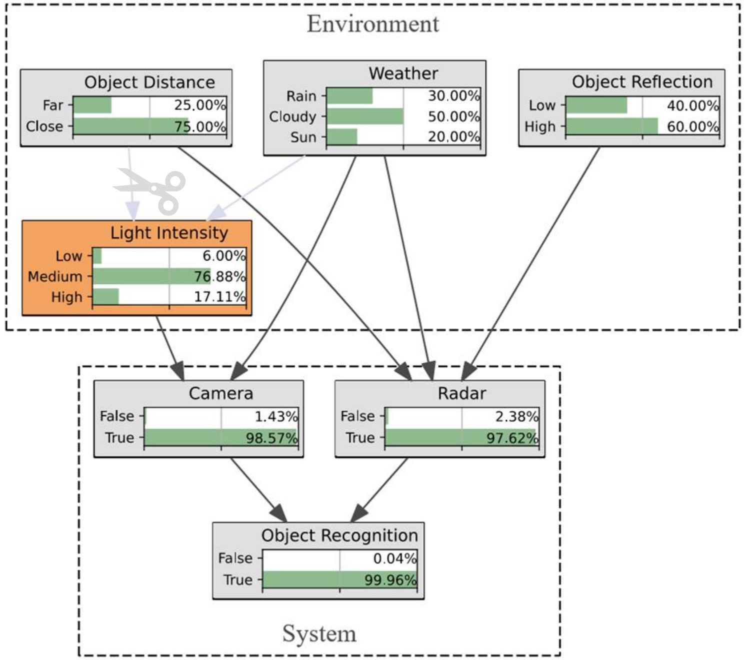
Investigation of Safety Measures



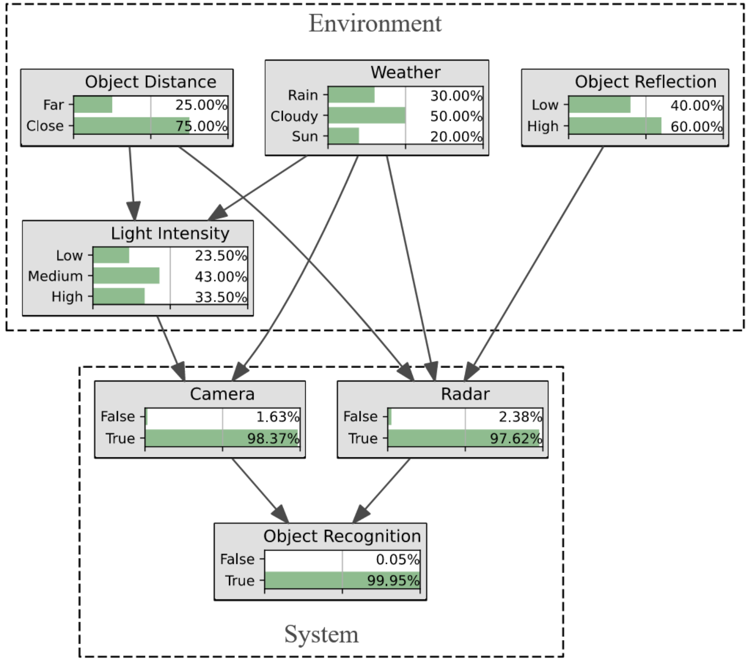
(1) Adaption of Context:



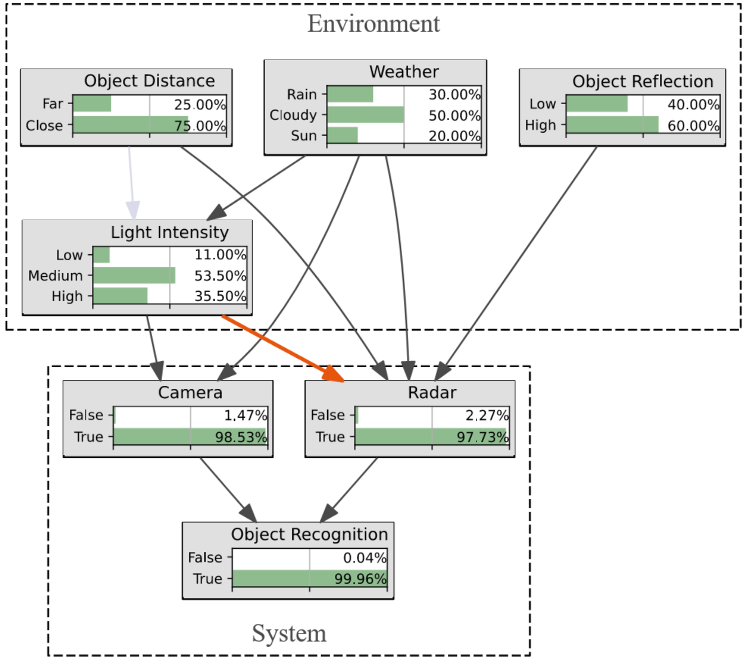
(2) (Stochastic) Interventions:



Investigation of Safety Measures



(3) Adaption of Causal Structure:



Future Work

- Application to real world use cases
- Causal modeling of dynamic interactions
- Integration into a model-based approach
- Modularity of causal models
- Application of causal learning
- Verification of causal assumptions on interventional level

Thank you for the attention.

Contact:

Lina Putze, M.Sc.
German Aerospace Center (DLR) e.V.
Institute of Systems Engineering for Future Mobility
lina.putze@dlr.de

Literature

[Pea09] Pearl, J., ’Causal Inference in Statistics: An Overview’, in Statistics Surveys, 2009

[Pe17] Peters et al., ‘Elements of Causal Inference: Foundations and Learning Algorithms’,2017, in the MIT Press.

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

[Ma23] Maier et al. ‘Causal Models to Support Scenario-Based Testing of ADAS,’ in IEEE Transactions on Intelligent Transportation Systems, 2024

[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.

[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.

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