Causality and Functional Safety - How Causal Models Relate to the Automotive Standards ISO 26262, ISO/PAS 21448, and UL 4600

Robert Maier ✉
Laboratory for Safe and Secure Systems (LaS³)
Regensburg University of Applied Sciences
Regensburg, Germany
robert.maier@oth-regensburg.de

Jürgen Mottok
Laboratory for Safe and Secure Systems (LaS³)
Regensburg University of Applied Sciences
Regensburg, Germany
juergen.mottok@oth-regensburg.de

2022 International Conference on Applied Electronics (AE) | 978-1-6654-9482-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/AE54730.2022.9920053

***Abstract*— With autonomous driving, the system complexity of vehicles will increase drastically. This requires new approaches to ensure system safety. Looking at standards like ISO 26262 or ISO/PAS 21448 and their suggested methodologies, an increasing trend in the recent literature can be noticed to incorporate uncertainty. Often this is done by using Bayesian Networks as a framework to enable probabilistic reasoning. These models can also be used to represent causal relationships. Many publications claim to model cause-effect relations, yet rarely give a formal introduction of the implications and resulting possibilities such an approach may have. This paper aims to link the domains of causal reasoning and automotive system safety by investigating relations between causal models and approaches like FMEA, FTA, or GSN. First, the famous "Ladder of Causation" and its implications on causality are reviewed. Next, we give an informal overview of common hazard and reliability analysis techniques and associate them with probabilistic models. Finally, we analyse a mixed-model methodology called Hybrid Causal Logic, extend its idea, and build the concept of a causal shell model of automotive system safety.**

Keywords*—*Causality*, *Functional Safety*, *Reliability

I. Introduction

Causality has become a trend to pursue in many fields. When one thinks about relating different variables and their interactions to each other, the famous phrase "correlation does not imply causation" comes to mind. In the context of functional safety (FS) and reliability of sociotechnical systems, causality has played an inherent role as a guiding paradigm. Standards like ISO 26262, which outlines and defines FS in the automotive use-case [1], or ISO/PAS 21448 which deals with the Safety Of The Intended Functionality (SOTIF) [2] formalize this. They do so by focusing on the malfunctioning behaviour of items as causes of hazardous events, which develop through causal mechanisms into hazards and eventually lead to harm.

Bayesian Networks (BNs) can serve as an efficient and flexible framework for probabilistic reasoning [3]. These models are used throughout many domains, including FS and reliability [4]–[7]. Many such publications claim (explicitly or indirectly) that their models are causal, yet rarely give a formal definition of causality or which implications a causal model may have. Additionally, existing work primarily treats automotive standards and their respective methodologies as independent frameworks. By extending these approaches to support reasoning under uncertainty, it becomes feasible to harmonize and combine them. Yet, a formal discussion summarizing notable foundational work regarding causal BNs and standard-compliant methodologies is missing. To address this demand, this paper aims to link automotive system safety with model-based causality. Its key contributions are an overview of basic research in this area, an investigation of causation, its individual implementation in different methodologies, and a generic model that conceptually merges standards and approaches. Moreover, the following questions will be addressed:

RQ1 Can standard-encouraged methodologies be linked to causal, probabilistic models?
RQ2 How can this be done?
RQ3 What advantages emerge by using causal models?

There are multiple frameworks for dealing with causality like the Potential Outcome framework [8], graph-based approaches like Single World Intervention Graphs [9], and one of the most common frameworks called Structural Causal Model (SCM) [10].

SCMs and their constrained graphical representations (i.e. BNs) will be used exemplary throughout this paper due to their high degree of maturity, widespread usage across various domains, and their accessibility when represented as causal diagrams.

First, an informal overview of distinct levels of causal expressiveness is presented. After highlighting the relationship between SCMs and BNs we will give pointers on how to build these models.

Next, a variety of common, standard-encouraged methodologies like Fault Tree Analysis (FTA) or Generalized Stochastic Petri-Nets (GSPNs), and their relations with causality, are reviewed. Finally, the relatively new Hybrid Causal Logic (HCL) framework is used as a baseline technique to show the potential of using causal models to bridge multiple modelling scopes. While most of the standard-encouraged approaches focus on a specific aspect of FS, generating a joint view on system safety is usually out of scope. We outline that by using causal models, high flexibility as well as the ability to work with one universal framework to address various areas of common standards arises. The resulting, novel view on FS and reliability will be called the *causal shell model* of automotive system safety.

II. Causal Models

The following section focuses exemplary on SCMs as presented in [10].

A. Causal Bayesian Networks and Structural Causal Models

BNs model associational relationships between variables as a graph, and are used to represent a joint probability distribution in an efficient, factorized, and easily inferable way [3]. One of the most important properties called the *local* Markov condition [10, Theorem 1.2.7] states that the represented joint distribution may be factorized as:

$$P(\mathbf{X}) = \prod_i P(x_i|pa_i) \tag{1}$$

where the operator pa_i is read as "parents of" the node x_i. A network built on this property encodes assumptions about dependencies as well as (conditional) independencies between variables, which can be formalized by the concept of d-separation [11]. One of the main implications is the so-called Markov Equivalence Classes (MECs) [12]. They state that a joint distribution can (under certain constraints) be represented by different BNs. If edges are interpreted as cause-effect relations, the associated model is called a Causal Bayesian Network (CBN).

Graph-based approaches to causality are well established and their formalisms, as well as algorithms for inference (computing probabilistic information), are based on proofed theorems [9], [10]. They can be used to answer questions like "what if I do", "what would have happened if", and "why". A famous interpretation of three distinct levels of causal expressiveness is given by [13] known as the "Ladder of Causation" and formally analysed as Pearl's Causal Hierarchy (PCH), for example by [14].

At the lowest level ($\mathcal{L}_1$) BNs can be used to encode associational relationships. On the second level ($\mathcal{L}_2$) CBNs are used as a framework to deal with interventional and counterfactual distributions. Structural Causal Models (SCMs) are suitable for the highest level ($\mathcal{L}_3$) of causative expressiveness and allow in principle the representation of all causal relationships, as well as the ability to answer any causal query. As [14] points out, higher-level representations usually entail lower levels.

SCMs are a framework to model causal processes (e.g. functional relationships) that assign probability distributions to a variable based on its influences. These mechanisms are independent of each other, can be locally replaced or modified, and allow modularity.

A formal definition of SCMs following [10, p. 203] and [14] can be given as:

Definition 1 (Structural Causal Model (SCM) [14])**.** *A SCM is a 4-tuple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F}, P(u) \rangle$ where*

1) $\mathbf{U}$ *is a set of background variables (also called exogenous) that are determined by factors outside the model.*
2) $\mathbf{V} = \{V_1, ..., V_n\}$ *is a set of endogenous variables that are determined by variables in the model, viz. variables in* $\mathbf{U} \cup \mathbf{V}$.
3) $\mathbf{F}$ *is set of functions* $\{f_1, ..., f_n\}$ *such that each* f_i *is a mapping from (the respective domains of)* $U_i \cup PA_i$ *to* V_i*, where* $U_i \subseteq \mathbf{U}$ *and* $PA_i \subseteq \mathbf{V} \setminus V_i$ *and the entire set of* $\mathbf{F}$ *forms a mapping form* $\mathbf{U}$ *to* $\mathbf{V}$*. In other words,* f_i *assigns a value to the corresponding* $V_i \in \mathbf{V}, v_i \leftarrow f_i(pa_i, u_i)$*, for* $i = 1, ...n$.
4) $P(u)$ *is a probability function defined over the domain of* $\mathbf{U}$.

In this paper, causal models are interpreted as frameworks that allow causal reasoning [10]. Causal models formalize how probability distributions change under observations or external interventions.

B. Building Models

Approaches for building a model can be broadly categorized as:

Model transformation: A causal model can be created by adapting an existing model (domain adaption) or by translating established methods like FTA, Failure Mode and Effects Analysis (FMEA), and others into a BN. Research in the field of FS and reliability mostly focuses on model-to-model transformations.

Expert elicitation: Models can be built from human knowledge by interviewing domain experts. How elicitations can be conducted, or how deviating individual estimates may be combined, is an active field of research [15].

Data-driven: If suitable data can be provided, algorithms can be applied to learn networks up to their MEC [16]. There are also approaches for estimating the causal direction between two variables, as well as learning a full SCM [17]. Data-driven methods can be combined with expert-based processes to form an iterative approach.

Algorithmically generated models typically face the problem of MECs. These models are nonetheless well suited to answer associational queries on $\mathcal{L}_1$ of PCH and are commonly known as BNs. Using them as causal models usually requires human experts to provide the necessary causal directions of influence (i.e. specifying what is cause and effect).

An emerging trend in the FS and reliability community is to convert or link approaches like FTA, or Event Tree Analysis (ETA) to BNs. This allows modelling of common cause effects, incorporating environmental influences, and improving the expressiveness of some established methods, yielding additional modelling capabilities [4], [7]. These approaches usually generate PCH $\mathcal{L}_1$ and $\mathcal{L}_2$ models but are almost always used to answer associational queries, even though actual causal research could be conducted.

III. Causality and Modern Safety

Established standard-encouraged methods like Reliability Block Diagram (RBD), FTA, ETA, or FMEA and their many variations deal with cause-effect relationships. These methods cast logical and causal dependencies into an appropriate framework, and in many cases use probability theory for a quantitative description of component behaviour. Depending on the complexity, or the level of abstraction, the dynamic evolution of a system is treated as well. Even though causation is a key idea, there are still nuances as to how causality is finally incorporated.

A. Modelling Intentions of Common Approaches

As a graphical and logical method, ETA starts at a potential cause of various failure chains. Each event path between a trigger and a final system state (i.e. consequence) describes a logical failure propagation throughout the inspected system. Branching probabilities are specified, which allow the calculation of likelihoods for the different outcomes. Due to the requirement of mutually exclusive individual events and consideration of one trigger at a time, a stringent causal analysis of a system, may be problematic. In contrast, FMEA tries to systematically investigate single potential causes of component failure and their direct effects on an analysed system. Due to the focus on one failure at a time, an analysis of complex interactions between multiple failures is not readily available.

As a deductive method, FTA uses basic events that in logical, boolean combinations lead to an undesired top-level event (TLE). An FTA can serve as a qualitative and a quantitative tool, which can consider the degradation of components over time.

Markov-Models (MMs) allow specifying dynamic, probabilistic system behaviour. They are a graphical representation of memoryless systems consisting of different logically connected states. Transitions between states are stochastic and can be used to consider failure and repair intervals. Similarly, Generalized Stochastic Petri-Nets (GSPNs) can be used to model complex system behaviour on a quantitative and qualitative level. Compared to MMs, GSPNs bind transitions to requirements (i.e. firing conditions). System state configurations called markings can be used to configure an initial situation, and throughout a system's simulation may be used to define reachability sets. Additionally, controlling and logical actions can be used to provide further levels of abstraction.

Based on the complexity and often times cross-domain composition of systems (i.e. mix of hardware, software, and machine learning (ML) methods) traditional techniques, as listed above, may not be enough to describe correct system performance. Instead, methods like Systems Theoretic Process Analysis (STPA) treat the intended behaviour of a system as an additional requirement. STPA views a sociotechnical system as a process and control problem. Starting from unsafe control actions, triggering events and scenarios are identified which may cause hazardous system states and subsequently a loss (e.g. of human life) [18].

Traditional hazard analysis frameworks decompose a system into logical elements, which individually or in a chain of events contribute to failures. In the scope of ISO/PAS 21448 potential hazardous situations are a result of (intended) unsafe interaction between working (i.e. within their requirements) components.

System safety can also be viewed as a property that is achieved by fulfilling e.g. functional, operational, or autonomy safety goals. Goals are usually the results of abstract, technology-agnostic safety cases. They are the main aspect of standards like UL 4600 [19] and are often modelled by frameworks for managing argumentations like Claims Arguments Evidence (CAE) or Goal Structuring Notation (GSN). GSN provides rules and symbols to describe a series of statements, forming a graphical notation called "goal structure". Claims about a system can be fulfilled by the available evidence (solutions) [20].

As [21] outlines, many methodologies fail to address some desired FS requirements individually and a combination of techniques should be used instead. One relatively new method is the Hybrid Causal Logic (HCL) framework [22]. HCL combines Event Sequence Diagram (ESD), FTA, and BNs in a joint manner to create a holistic approach that can address goals, FS, SOTIF aspects, and complex environmental influences in a quantitative and qualitative way. As [21], [22], and others point out, one of the main advantages is the usage of a Probabilistic Graphical Model (PGM) (here a BN), that extends modelling capabilities like handling common cause influences while remaining transparent and computationally efficient.

B. Functional Safety Frameworks and Their Relationships With Causal Models

In the last two decades, various attempts have been made to enhance methods like ETA by relating or transforming them into BNs. One of the most common transformations applies to Fault Trees. While modelling common cause failures is usually problematic, doing so in a BN becomes trivial. In the case of FTA, using BNs can even be considered an upgrade [4].

The main motivation to use these models, besides the technical advantages (e.g. efficient factorization and inference of large models) is the ability to reason under uncertainty. PGMs enable us to encode knowledge about how things (causally) interact, summarize assumptions (e.g. technical simplifications), and allow tractable, transparent, and quantifiable judgment about a modelled system [3], [23].

Model transformations are common [24]–[28], and linking PGMs to FS and reliability methods recently became an active field of research. Connections between different variants of BNs and FMEA [29], RBD [6], FTA [4], [30], ETA [7], STPA [31], GSN [32], and others have been discussed. [33] and [34] show, that Dynamic Bayesian Networks (DBNs) can be used to represent some MMs. [35] shows how SCMs can be related to ordinary differential equations.

Figure 1 provides an informal, exemplary overview of conducted research on linking various standard-compliant methods.

As mentioned in section I, many publications in FS and reliability research use *causal* BNs, yet rarely give a definition of what *causal* means in this context. One of the implications of a non-causal network is MECs. As long as associational $\mathcal{L}_1$-queries like $P(X|Y,W)$ are concerned, these networks may provide correct results - yet fail to give correct answers for $\mathcal{L}_2$-queries like $P(X|\mathbf{do}(Y),W)$.

The main advantage which, comes from using causal models, is that of causal reasoning and interventions ($\mathbf{do}(X)$) in particular. Interventions allow modifying causal mechanisms and estimating these effects for relevant variables. Causal questions (e.g. "did component X cause the failure?") can be phrased as interventions ($\mathbf{do}(X = not\_working)$). Given

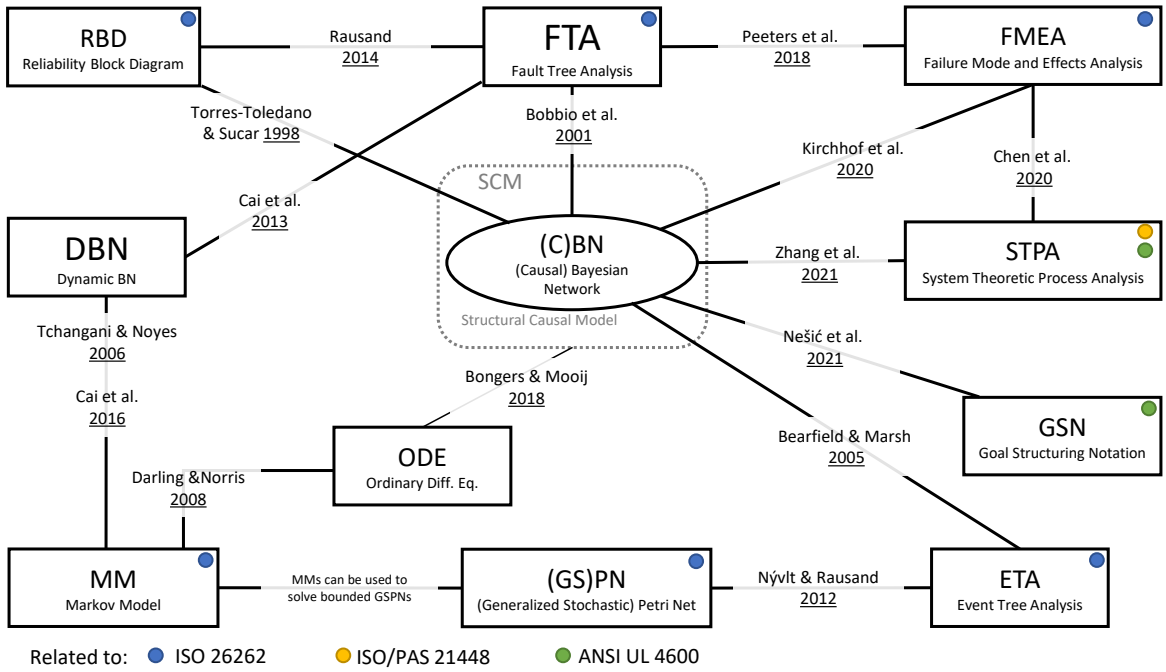

Fig. 1. Connections between various standard-encouraged analysis methods. Edge markings show exemplary work, linking the different approaches.

that such an ($\mathcal{L}_2$) query is identifiable [10, Definition 3.2.3], the effects of interventions can be calculated based on available knowledge that is encoded in the causal model. This can be done without the need to generate or collect new data, which is compliant with the modified system. On top of this, counterfactual queries (e.g. "What would have happened if component X did not fail") can be estimated.

Interventions ("doing") and counterfactuals ("imagining") are fundamentally different from observations ("seeing"). Due to the modification of causal mechanisms, the probability distributions that are described and entailed by an underlying SCM change. Compared to the associational case, this affects the estimate of a query (if identifiable) based on the causal structure of the model (see also [10], [14] for a detailed discussion).

As mentioned in section III-A the scope of FS, decomposition and logical (but not necessarily causal) combination of sub-systems is a standard way to manage system complexity. Related analysis techniques are rather linear, in the sense that they treat chains of connected events or item interactions as a causal chain. Decomposition and individual treatment of components become gradually problematic when trying to evaluate the intended functionality of a system (i.e. SOTIF). Additionally, modelling and quantifying the effects of environmental conditions (as they are present in scenarios) is outside the scope of most methodologies. Hybrid approaches like HCL try to compensate for the increasing complexity in parts by combining different, suitable frameworks. These focus on distinct levels of abstraction (i.e. environmental and technical aspects, dynamic aspects, or required goals). Even though they are able to cover up many downsides of the individual approaches used [21], causal reasoning as mentioned above is hardly feasible.

As [21] points out, the high flexibility of HCL is in large parts driven by the usage of a BN that can represent environmental or common cause influences. Although HCL distinguishes between three layers (i.e. BN, FTA, ESD), these may in practice very well be modelled as one joint BN or ideally as an SCM. This would allow using one unifying framework to address a joint multi-model approach.

C. High-level View on System Safety

In the automotive industry, established standards like ISO 26262 and its relative ISO/PAS 21448 shape how system safety is addressed. Due to the rapid development of ML-based components, systems using such technologies are not adequately covered by the above-mentioned standards any more [36]. Therefore, new approaches like safety case argumentations are required, which tend to be independent of a specific system and instead goal-oriented. One relatively new standardization approach is UL 4600 [19], which focuses on the usage of methodologies like GSN, among others.

From the perspective of causal modelling, automotive system safety can be split into different domains, where each is covered by one of the above standards. An item under test typically consists of multiple, interconnected, and diverse components, that together constitute a system with intended functionality. The environment (nature, scenarios, or operational context) among others, may negatively influence such a system. Generic safety goals for item behaviour can be defined alongside technical component specifications. Designed item performance may be evaluated by scenario-based testing (as outlined in ISO/PAS 21448), hazard and reliability analysis approaches (ISO 26262), or by fulfilling valid safety cases (UL 4600). In this context, the HCL framework may be viewed as a methodology that tries to tie the domains of these three standards together [21].

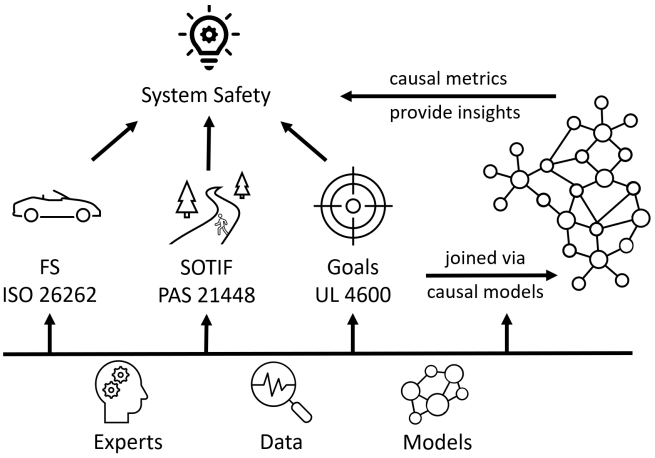

Fig. 2. Causal models (e.g. SCMs) can be used to transform and link established standard-encouraged methodologies. Causal knowledge in the form of (abstract) environmental or situational influences, expert knowledge, or data can be incorporated. Causal insights are generated and allow a joint 3+1 view of automotive system safety.

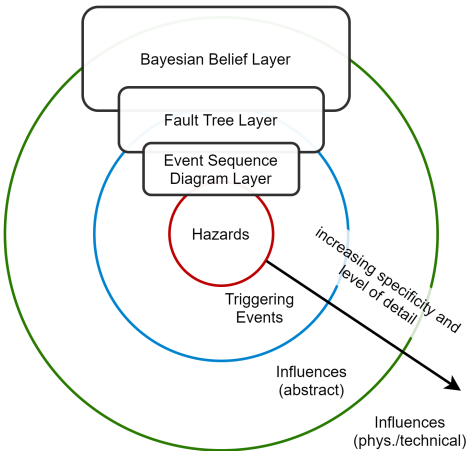

Fig. 3. The causal shell model consists of three intersecting layers that relate to different perspectives of system safety. All layers may be connected using causal models (e.g. CBNs). Exemplary, the layers of HCL (rectangles) are included to show their modelling scopes.

As mentioned in section III-B, many standard-encouraged techniques can be transformed into causal models. This allows causal reasoning, which can be applied to generate causal insights. The underlying models serve as containers for causal knowledge and may either be generated by human experts, model transformation, (partly) from data or as an iterative process combining the above. Instead of treating aspects of decomposed components and events individually, merged causal models allow to jointly consider system safety. This can be achieved by connecting individual approaches via adequate causal mechanisms serving as links.

It should be noted that a joint consideration is not always necessary or feasible. Depending on the underlying methodologies used (e.g. FTA or ETA) some combinations may not be reasonable and should therefore be treated independently. Figure 2 shows a high-level view of these interactions, where the combination of three domains (system, behaviour, and goals) leads to an encompassing view of system safety based on causal models.

Another perspective on viewing system safety is that of avoiding accidents that are the results of unhandled hazards. Common cause failures, environmental conditions, situational effects like road accidents, or other causes like an inattentive or stressed human driver, may serve as triggers for hazardous events. If unhandled, they lead to harm or loss as the result of distinct causal mechanisms.

Depending on which end an analysis starts (top-down or bottom-up), different layers of abstraction can be identified. At the core, potential hazards are considered. Based on the required level of abstraction, triggering events can be detailed on a social, technical, or physical level. By specifying the causal processes that lead to an event, influencing factors can directly or indirectly contribute to a hazardous system state (e.g. basis events and their deterministic combination via boolean gates). Each contributing factor and mechanism may be identified by the approaches discussed above. To achieve a comprehensive view of system safety, different perspectives (i.e. scopes of the standards) can be combined end-to-end in a joint causal model.

Figure 3 shows the conceptual idea of our proposed *causal shell model*. As an example of an appropriate combination of analysis techniques, the layers of the HCL framework are included, showing the coverage of all shell levels.

IV. Conclusion and Future Work

Many established methods suggested by relevant automotive safety standards can be adapted to allow reasoning under uncertainty. This can be done by using CBNs (or ideally SCMs) as a suitable framework to model all levels of PCH (RQ1 & RQ2). The main contribution of this paper is that of a *3+1 view of automotive system safety*, consisting of three relevant standards and causal models as a framework to jointly consider them. Based on this premise, the conceptual idea of a *causal shell model* is proposed. It builds on the results of [21] and extends its idea of using different but suitable analysis methodologies to cover the various scopes of the above standards. Instead of resorting to diverse frameworks like in HCL, using causal models as a domain-independent, universal, and highly versatile approach should be encouraged. This allows accessing data-driven methods and expert knowledge to build models.

Causal frameworks allow connecting multi-domain models and different levels of abstraction. This renders causal reasoning a central methodology that can address and complement many desired FS and reliability requirements (RQ3).

In the context of scenario-based testing, the influence of the environment is of central importance. To evaluate the safety of a scenario, it is necessary to model and work with the operational design domain of the system under test. Although recent literature suggests using the methods presented in this paper, no proper tool support is available. To help investigate the implications of using causal models, we plan to provide an Open-Source software package called *BayesianSafety*[1]. It will allow evaluating environmental influences on FTA and ETA, as well as

[1]https://github.com/othr-las3/bayesiansafety

researching the outlined approach for automotive system safety. The main focus of future research will be on how to create expressive causal models. Engineering them should be possible in conjunction with established industrial processes. Ideally, existing methodologies like FMEA can act as a starting point to build models and evaluate insights derived from them. Additionally, measures, metrics, and thresholds need to be defined, as they are a necessary element for using the *causal shell model*.

ACKNOWLEDGMENT

The present paper is supported by *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie* through the granting of the funding project *HolmeS³* (FKZ: DIK0173/03). We thank L. Grabinger and D. Urlhart for valuable discussions.

REFERENCES

[1] ISO Central Secretary, "Road vehicles – functional safety part 2: Management of functional safety," International Organization for Standardization, Geneva, CH, Standard ISO 26262-2:2018, 2018.

[2] ——, "Road vehicles – safety of the intended functionality," International Organization for Standardization, Geneva, CH, Standard ISO/PAS 21448:2019, 2019.

[3] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*, ser. Adaptive computation and machine learning. Cambridge, MA: MIT Press, 2009.

[4] A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla, "Improving the analysis of dependable systems by mapping fault trees into Bayesian networks," *Reliability Engineering & System Safety*, vol. 71, no. 3, pp. 249–260, March 2001.

[5] S. Kabir and Y. Papadopoulos, "Applications of bayesian networks and petri nets in safety, reliability, and risk assessments: A review," *Safety Science*, vol. 115, pp. 154–175, February 2019.

[6] J. Torres-Toledano and L. Sucar, *Bayesian Networks for Reliability Analysis of Complex Systems*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, January 1998, vol. 1484.

[7] G. Bearfield and W. Marsh, "Generalising Event Trees Using Bayesian Networks with a Case Study of Train Derailment," in *Computer Safety, Reliability, and Security*, ser. Lecture Notes in Computer Science, R. Winther, B. A. Gran, and G. Dahll, Eds. Berlin, Heidelberg: Springer, 2005, pp. 52–66.

[8] D. B. Rubin, "Causal inference using potential outcomes," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 322–331, 2005.

[9] T. S. Richardson and J. M. Robins, "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality," in *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, vol. 30, no. 128. Citeseer, 2013.

[10] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. USA: Cambridge University Press, 2009.

[11] D. Geiger, T. Verma, and J. Pearl, "d-Separation: From Theorems to Algorithms," in *Machine Intelligence and Pattern Recognition*. Elsevier Science Inc., 1990, vol. 10, pp. 139–148.

[12] T. Verma and J. Pearl, "Equivalence and synthesis of causal models," in *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, ser. UAI '90. USA: Elsevier Science Inc., 1990, p. 255–270.

[13] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, 1st ed. USA: Basic Books, Inc., 2018.

[14] E. Bareinboim, J. D. Correa, D. Ibeling, and T. F. Icard, "1 On Pearl ' s Hierarchy and the Foundations of Causal Inference," Columbia University, Technical report 60, 2021.

[15] E. Nyberg, A. Nicholson, K. Korb, M. Wybrow, I. Zukerman, S. Mascaro, S. Thakur, A. Oshni Alvandi, J. Riley, R. Pearson, S. Morris, M. Herrmann, A. K. M. Azad, F. Bolger, U. Hahn, and D. Lagnado, "BARD: A structured technique for group elicitation of bayesian networks to support analytic reasoning," *Risk Analysis*, June 2021.

[16] M. J. Vowels, N. C. Camgöz, and R. Bowden, "D'ya like DAGs? A survey on structure learning and causal discovery," *CoRR*, vol. abs/2103.02582, 2021. [Online]. Available: https://arxiv.org/abs/2103.02582

[17] Y. Yang, M. Nafea, A. Ghassami, and N. Kiyavash, "Causal Discovery in Linear Structural Causal Models with Deterministic Relations," *arXiv:2111.00341 [cs, math, stat]*, October 2021, arXiv: 2111.00341.

[18] N. Leveson and J. Thomas, "STPA handbook," 2018.

[19] Underwriters Laboratories Inc., "Standard for evaluation of autonomous products," Underwriters Laboratories Inc., Standard ANSI/UL 4600:2020, 2020.

[20] S. A. C. W. Group, "Goal structuring notation community standard (version 3), may 2021," SCSC Assurance Case Working Group, Standard SCSC – 141C, 2021.

[21] S. Thomas and K. Groth, "Toward a hybrid causal framework for autonomous vehicle safety analysis," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, p. n. a., August 2021.

[22] A. Mosleh, A. Dias, G. Eghbali, and K. Fazen, "An integrated framework for identification, classification, and assessment of aviation systems hazards," in *Probabilistic Safety Assessment and Management*, C. Spitzer, U. Schmocker, and V. N. Dang, Eds. London: Springer London, 2004, pp. 2384–2390.

[23] J. Pearl, "Chapter 1 - Uncertainty in AI systems: an overview," in *Probabilistic Reasoning in Intelligent Systems*, J. Pearl, Ed. San Francisco (CA): Morgan Kaufmann, 1988, pp. 1–28.

[24] J. Peeters, R. Basten, and T. Tinga, "Improving failure analysis efficiency by combining FTA and FMEA in a recursive manner," *Reliability Engineering & System Safety*, vol. 172, pp. 36–44, 2018.

[25] L. Chen, J. Jiao, and T. Zhao, "A novel hazard analysis and risk assessment approach for road vehicle functional safety through integrating STPA with FMEA," *Applied Sciences*, vol. 10, no. 21, 2020.

[26] M. Rausand, *Reliability of Safety-Critical Systems: Theory and Applications*, 1st ed. Wiley Publishing, 2014.

[27] R. Darling and J. Norris, "Differential equation approximations for markov chains," *Probability Surveys*, vol. 5, January 2008.

[28] O. Nývlt and M. Rausand, "Dependencies in event trees analyzed by petri nets," *Reliability Engineering & System Safety*, vol. 104, pp. 45–57, 2012.

[29] M. Kirchhof, K. Haas, T. Kornas, S. Thiede, M. Hirz, and C. Herrmann, "Root Cause Analysis in Lithium-Ion Battery Production with FMEA-Based Large-Scale Bayesian Network," *arXiv:2006.03610 [stat]*, June 2020, arXiv: 2006.03610.

[30] B. Cai, Y. Liu, Y. Zhang, Q. Fan, Z. Liu, and X. Tian, "A dynamic Bayesian networks modeling of human factors on offshore blowouts," *Journal of Loss Prevention in the Process Industries*, vol. 26, no. 4, pp. 639–649, July 2013.

[31] S. Zhang, T. Tang, and J. Liu, "A Hazard Analysis Approach for the SOTIF in Intelligent Railway Driving Assistance Systems Using STPA and Complex Network," *Applied Sciences*, vol. 11, no. 16, p. 7714, August 2021.

[32] D. Nešić, M. Nyberg, and B. Gallina, "A probabilistic model of belief in safety cases," *Safety Science*, vol. 138, p. 105187, June 2021.

[33] A. Tchangani and D. Noyes, "Modeling dynamic reliability using dynamic Bayesian networks," *Journal Européen des Systèmes Automatisés*, vol. 40, October 2006.

[34] B. Cai, Y. Liu, and Q. Fan, "A multiphase dynamic Bayesian networks methodology for the determination of safety integrity levels," *Reliability Engineering & System Safety*, vol. 150, pp. 105–115, June 2016.

[35] S. Bongers and J. M. Mooij, "From Random Differential Equations to Structural Causal Models: the stochastic case," *arXiv:1803.08784 [cs, stat]*, March 2018, arXiv: 1803.08784.

[36] A. Rudolph, S. Voget, and J. Mottok, "A consistent safety case argumentation for artificial intelligence in safety related automotive systems," in *ERTS 2018*, ser. 9th European Congress on Embedded Real Time Software and Systems (ERTS 2018), Toulouse, France, January 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02156048