Structural and Sequential Cause in Hidden Markov Models

Katherine Braught
University of Illinois Urbana-Champaign
braught2@illinois.edu

Sayan Mitra
University of Illinois Urbana-Champaign
mitras@illinois.edu

Causal reasoning for sequential time-series observations is an important problem in a number of disciplines. Recently, Baier et al. have formalized the notion of cause of a language in Markov chains as prefixes that raise the probability of an effect. In this paper, we extend this notion of p-causes for Hidden Markov Models. A natural question is to investigate the relationship between this notion of p-causes and Pearl's notion of total causal effect defined using structural causal models and interventions. We explore this relationship for Hidden Markov Models. We find that even when we construct a SCM to have the same probability distribution as the HMM, the notions are incompatible: p-causes do not imply total causal effect and total causal effect does not imply a p-cause. This incompatibility comes from underlying conceptual differences. A p-cause is set of specific observation sequences that leads to a specific effect with high probability. A total causal effect in the constructed SCM implies that the current state at one time will effect the state at a later time.

1 Introduction

The potential for formal and causal analysis of software system has been noted in a number recent publications [3, 5, 8, 14, 15]. A number of different notions of cause and effect for sequential observations have been developed in the literature: The counterfactual principle [6] captures the idea that there is not effect without its cause. Granger causality is based on the cause preceding the effect in temporal observations [2]. Structural causal models (SCMs) [11] represent causal relationships among random variables as a graph and formalize the notions of confounding, interventions, and total causal effects [13]. SCMs have helped uncover unexpected causal relationships in domains as diverse as finance [9], vehicle accidents [7], and medicine [10]. Baier et al [4] have recently introduced the notion of *p-causes* for Markov chains formalizing the principle that a cause raises the probability for an effect. P-causes are sets of prefixes in a Markov chain that provide high probability of reaching an effect. Thus, p-causes capture both the probability raising principle and the counterfactual principle. In contrast to SCMs, p-causes consider the cause and effect of sequences of states, rather than the relationship between variables (regardless of their specific values). Despite this difference, is it possible for the two notions to coincide? In this paper, we begin to study how p-causes and SCMs apply to Hidden Markov Models.

HMMs are useful for studying sequential observations arising from stochastic processes with limited observability. How does the notion of p-causes apply to languages over observations instead of states? How can Hidden Markov Models be represented as structural causal models? Is there a shared meaning between p-causes and total causal effect? In this paper, we attempt to answer these questions via a particular type of unrolling of HMMs as an SCM. We call this the independent noise unrolling (IU) of an HMM. We find that both causality notions can provide insights on causal relationships of observation sequences over time. On the other hand, we find that, for our IU encoding of HMMs, these two concepts are not precisely comparable. There are examples where a p-cause will have a total causal effect in one HMM but not in another HMM, while both p-causes and non-p-causes can have total causal effects. From this study, we conclude that time series p-causes can be used to find what prefixes lead to a trace

Submitted to:
CREST 2023

language, and total causal effect helps understand the effect that different time steps have on each other. We present the main definitions and results and a complete version of the paper with proofs will be made available online [1].

2 Hidden Markov Models

Definition 2.1. A *Hidden Markov Model (HMM)* is a tuple $\mathbb{H} = \langle Q, S, P, O, \theta \rangle$, where Q is a finite set of *states*, S is finite set of *observations*, $P : Q \times Q \to [0,1]$ is a *probabilistic transition function*, $O : Q \times S \to [0,1]$ is a *probabilistic observation function*, and $\theta \in \mathbb{P}(Q)$ is an *initial distribution.*

A *execution* of an HMM $\mathbb{H}$ is a finite or infinite alternating sequence of states and observations $\alpha = (q_0, s_0), (q_1, s_1), \ldots$, such that (1), $\theta(q_0) > 0$, (2) $P(q_i, q_{i+1}) > 0$, (3) $O(q_i, s_i) > 0$, for each i in the sequence. The set of all executions is denoted by $Execs_{\mathscr{M}}$. Given a finite execution $\beta = (q_0, s_0), \ldots, (q_k, s_k)$, the *cylinder set of* β, denoted by $Cyl(\beta)$, is the set of all infinite executions starting with the prefix β. For an HMM $\mathbb{H}$, we define the probability of the cylinder set as

$$P_{\mathbb{H}}(Cyl(\beta)) = \theta(q_0)O(q_0, s_0) \times \prod_{i=1}^{k} P(q_{i-1}, q_i)O(q_i, s_i). \tag{1}$$

The probability distribution over cylinder sets $P_{\mathbb{H}}$ can be uniquely extended to a probability measure over the σ-algebra generated by all cylinder sets.

For a finite execution α, $\alpha \downarrow S$ is the corresponding observation sequence or *trace* and $\alpha \downarrow Q$ gives the corresponding state sequence or *path*. The probability distribution over a finite trace $\sigma = s_0, s_1, \ldots, s_k$ is defined using the distribution over executions:

$$P_{\mathbb{H}}(Cyl(\sigma)) = \sum_{\alpha : \alpha \downarrow S = \sigma} P_{\mathbb{H}}(Cyl(\alpha)). \tag{2}$$

The probability distribution over $Traces_{\mathbb{H}}$ is defined in the usual way. A finite trace σ defines a *cylinder set* which contains all possible extensions of σ: $Cyl(\sigma) = \{\beta \in S^{\omega} \mid \sigma \in Pref(\beta)\}$. The sets of all non-zero probability executions and traces are $Execs_{\mathbb{H}}$ and $Traces_{\mathbb{H}}$. We have overloaded $P_{\mathbb{H}}$ to denote both the probability distributions over traces or executions, but the arguments used will disambiguate.

Consider an ω-regular language $\mathscr{L} \subseteq S^{\omega}$. The probability $P_{\mathbb{H}}(\mathscr{L})$ is simply the sum of the probabilities of all the strings in $\mathscr{L}$. For a finite trace β with non-zero probability for the corresponding cylinder sets, the conditional probability of $\mathscr{L}$ given β is:

$$P_{\mathbb{H}}(\mathscr{L}|\beta) = \frac{P_{\mathbb{H}}(\mathscr{L} \cap Cyl(\beta))}{P_{\mathbb{H}}(Cyl(\beta))}. \tag{3}$$

For brevity we write $P_{\mathbb{H}}(\mathscr{L}|Cyl(\beta))$ as $P_{\mathbb{H}}(\mathscr{L}|\beta)$. Later, we use linear temporal logic to specify languages, $\mathscr{L}$. We use the eventually ($\diamond$) and next () operators defined in [1]

3 P-Cause in Hidden Markov Models

In this section, we introduce the notion of p-causes for trace languages of HMMs. A p-cause for a language $\mathscr{L}$ is a finite trace that makes the probability of observing a string in $\mathscr{L}$ larger than p. We follow

[1] `https://mitras.ece.illinois.edu/pubs.html`

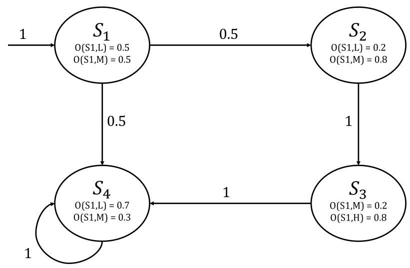


Figure 1: A HMM representing a system that can overheat

the developments in [4], where a similar notion was studied for Markov chains; however, an additional complexity arises here because the same trace could be generated by wildly different executions of a Hidden Markov Model. Throughout this section, we fix an HMM $\mathbb{H} = \langle Q, S, P, O, \theta \rangle$, an ω-regular language $\mathscr{L} \subseteq S^\omega$, and a constant $p \in (0,1]$.

First, we define p-critical prefixes of a language. A finite trace is p-critical if once this trace is observed, with at least probability p, all extension of the current execution will generate an execution with a trace in $\mathscr{L}$.

Definition 3.1. For any HMM $\mathbb{H}$, any nonempty language $\mathscr{L} \subseteq S^\omega$ and any $p > 0$, a finite trace $\sigma \in S^*$ is a *p-critical prefix* if $P_{\mathbb{H}}(\mathscr{L}|\sigma) \geq p$.

A p-critical prefix can be useful to know, at a finite point in time, that we are likely observe a string in the target language in the future. If the language encodes reaching an unsafe state or exiting a safe state, then the p-critical prefixes could be used for monitoring.

Example 1. Consider a machine with (hidden) states $Q = \{S_1, S_2, S_3, S_4\}$ as shown in Figure 1. At each state the observed temperature is $S = \{L, M, H\}$ (corresponding to low, medium, and high readings) with probabilities as shown. Probabilities not shown are considered probability zero. Suppose we are interested in the trace language: eventually H , that is, $\mathscr{L} = \diamond H$. Finding a p-critical prefix will help find executions where we are likely to observe high heat. Consider the prefix trace $\gamma = MM$. With explicit knowledge of the states and transitions, here we can easily calculate $P_{\mathbb{H}}(\mathscr{L}|\gamma) = 0.58$. We calculate $P_{\mathbb{H}}(L \cap Cyl(\gamma))$ by adding the probabilities of all paths that start with γ and are in $\mathscr{L}$, and we calculate $P_{\mathbb{H}}(Cyl(\gamma))$ by adding the probabilities of all paths that start with γ. Since $P_{\mathbb{H}}(\mathscr{L}|\gamma) \geq 0.5$, we can say that (MM) is a 0.5-critical prefix of $\mathscr{L} = \{\diamond H\}$.

These p-critical prefixes can be used for monitoring purposes. In the example above, once a MM is observed during an execution, there is a high probability of overheating, so a user may shut down the machine instead of allowing it to overheat. This particular prefix forewarns of a likely overheating, and therefore, is a useful trace to monitor.

A simple but useful property of p-critical prefixes is how they perform in subsets.

***Proposition* 1.** For any non-empty languages, $\mathscr{L}, \mathscr{L}' \subseteq S^\omega$, if γ is a p-critical prefix for $\mathscr{L}$ and $\mathscr{L} \subseteq \mathscr{L}'$, then γ is also a p-critical prefix for $\mathscr{L}'$.

Observing a p-critical prefix guarantees an increase in the probability of observing a string the language. However, a single p-critical prefix may not be a prefix for every trace in a language $\mathscr{L}$. Therefore, in Definition 3.2, we introduce p-causes which are a set of p-critical prefixes where every trace is $\mathscr{L}$ has a prefix in the p-cause.

Definition 3.2. For an HMM $\mathbb{H}$, a *p-cause* for $\mathscr{L} \subseteq S^\omega$ is a prefix free set of (non-zero probability) finite traces $\Pi \subseteq S^*$ such that:

1. Every $\pi \in Traces_{\mathbb{H}} \cap \mathscr{L}$ has a prefix $\hat{\pi} \in \Pi$.
2. Every $\hat{\pi} \in \Pi$ is a p-critical prefix for $\mathscr{L}$.

The two conditions for p-cause in Definition 3.2 correspond to the probability raising property and the counterfactual principle for cause and effect. Condition 1 guarantees that the 'effect' (a trace being in $\mathscr{L}$) would not have happened without the 'cause' (observing a trace in the p-cause). Condition 2 guarantees that the probability of the 'effect' increases after observing the 'cause'.

In the following example, we see that probability of being in $\mathscr{L}$ raises when we observe a prefix in the p-cause and a trace cannot be in $\mathscr{L}$ without first observing one of the prefixes in the p-cause.

Example 2. To find a p-cause for the HMM $\mathbb{H}$ in Example 1 and the language $\mathscr{L} = \diamond H$, we first find a p-critical prefix for every trace in $\mathscr{L}$. The cylinder sets of traces of the form $(L+M)(L+M)H$ represent all traces in $\mathscr{L}$. We showed in Example 1 that MM is a 0.5-critical prefix so it can be in the p-cause that covers $Cyl(MMH)$.
Consider the cylinder set of the remaining traces in $Traces_{\mathbb{H}} \cap \mathscr{L}$: $(LMH), (LLH), (MLH)$. As in example 1, we can find that $P(\mathscr{L}|LM) = .58$ and $P(\mathscr{L}|ML) = P(\mathscr{L}|LL) = 0.18$. Since LM is a .5-critical prefix, the prefix LM should be added to the p-cause to cover traces in $Cyl(LMH)$. To find a prefix of $Cyl(LLH)$ and $Cyl(MLH)$ that is $.5-$critical, LL and ML do not have high enough conditional probabilities. Therefore, LLH and MLH must be in the p-cause set. Therefore, a valid $\frac{1}{2}$-cause is $\{(MM), (LM), (LLH), (MLH)\}$

It is straightforward to see that p-causes are closed under union, but not under intersection.

***Proposition* 2.** For any non-empty languages, $\mathscr{L}, \mathscr{L}' \subseteq S^{\omega}$ and some HMM $\mathbb{H}$, if $\Pi \subseteq S^*$ is a p-cause for $\mathscr{L}$ and $\Pi' \subseteq S^*$ is a p-cause for $\mathscr{L}'$, then $\Pi \cup \Pi' \subseteq S^*$ is a p-cause for the language $\mathscr{L} \cup \mathscr{L}'$.

Lastly, we claim that there exists a p-cause for any p. In future work, we provide a method to transform all languages to a reachability language which will prove Proposition 3.

***Proposition* 3.** For any non-empty language $\mathscr{L} \subseteq S^{\omega}$, there exists a p-cause for any $p \in (0,1]$.

4 Structural Causal Models

Structural causal models (SCM) are used to represent and discover causal relationship between variables using the representation of a graph [11, 13]. SCMs are used for studying causal relationships between variables, interventions, and counterfactuals and these different levels of reasoning are organized in Pearl's hierarchy [12]. Most importantly, they can be used to find confounding factors in relationships between variables that appear directly related.

SCMs do not inherently have a notion of temporal evolution. However, different authors have used SCMs to reason about time-series data and dynamical systems using variables to explicitly represent the unrolling of the dynamic process.

4.1 Preliminaries

We provide a brief overview of Structural Causal Models using definitions from Peter's Elements of Causal Inferences [13].

Definition 4.1. A *structural causal model (SCM)* $\mathbb{C} = \langle \mathbf{X}, \mathbf{N}, \mathbf{E}, P_N \rangle$ consists of random variables $\mathbf{X}$, noise variables $\mathbf{N}$, and a collection $\mathbf{E}$ of d structural assignments of the form

$$X_j := f_j(PA_j, N_j), \tag{4}$$

where $PA_j \subseteq \{X_1, \ldots, X_d\}$ are called *parents* of X_j; and P_N is a joint distribution over the noise variables $N_1, \ldots, N_d$. The noise variables are jointly independent and the dependency graph over X_j's is acyclic.

Given an SCM, an *intervention* on a variable is a change to the structural assignment of that variable. It can change the distribution of the system so that the variable is not influenced by any of its parents in the original SCM.

Definition 4.2. Given an SCM $\mathbb{C} := \langle \mathbf{X}, \mathbf{N}, \mathbf{E}, P_N \rangle$ and its entailed distribution $P_{\mathbf{X}}^{\mathbb{C}}$, an *atomic intervention* replaces one (or several) structural assignments to obtain a new SCM $\tilde{\mathbb{C}}$. The intervention replaces the assignment for X_k by

$$X_k := a$$

where a is some constant value in the range of X_k. The entailed distribution of $\tilde{\mathbb{C}}$ is the *intervention distribution* and the variables with new structural assignments have been *intervened* on. The new distribution is denoted

$$P_{\mathbf{X}}^{\tilde{\mathbb{C}}} = P_{\mathbf{X}}^{\mathbb{C};do(X_k:=a)}.$$

Interventions can be used to study the causal relationship between variables. In our study of p-cause, we will specifically use total causal effect. Total causal effect from X to Y implies that to X has an effect on Y either directly or indirectly (through other vertices) and can be checked using Equation 5.

***Proposition* 4.** Given an SCM $\mathbb{C} = \langle \mathbf{V}, \mathbf{N}, \mathbf{E}, P_N \rangle$, there is a *total causal effect* from $X \subseteq \mathbf{V}$ to $Y \subseteq \mathbf{V}$ iff there is an assignment a that gives a constant value for each $X_i \in X$ such that

$$P_Y^{\mathbb{C};do(X:=a)} \neq P_Y^{\mathbb{C}}. \tag{5}$$

To compute $P_Y^{\mathbb{C};do(X:=a)}$, we use *adjustment sets*, which are typically sets of other vertices that can have an effect on X and Y and allow us to calculate interventions using conditional probabilities.

All of this is implicitly in terms of real-valued random variables. Our construction later will be in terms of bit vectors to represent discrete values.

4.2 Hidden Markov Models as Structural Causal Models

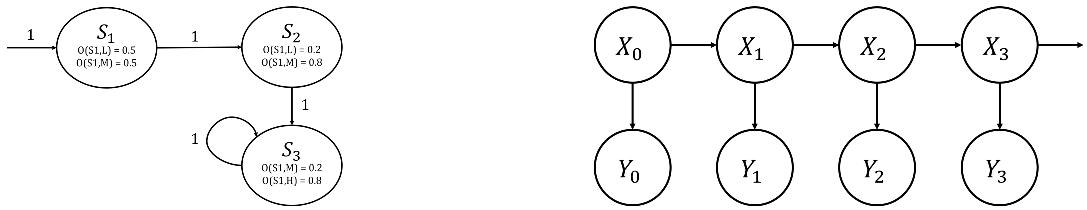
Now, we can construct a SCM that entails a probability distribution that is equivalent to the probability distribution of an HMM's states and observations over T time steps. This SCM is called the independent noise unrolling of an HMM ($IU(\mathbb{H})$). Time series data SCMs are presented in [13]; however, we present the specific equations needed to represent the probability distribution of an HMM. Given Definition 4.3, we claim that for any execution $\gamma = x_0, y_0, \ldots x_T, y_T$ of $\mathbb{H}$, $P_{\mathbb{H}}(Cyl(\gamma)) = P^{\mathbb{C}}(X_0 = x_0 \wedge Y_0 = y_0 \wedge \ldots \wedge X_T = x_T \wedge Y_T = y_T)$.

Definition 4.3. The SCM $\mathbb{C} = IU(\mathbb{H})$ for the *independent noise unrolling* of an HMM $\mathbb{H} = \langle Q, S, P, O, \theta \rangle$ for any natural number T, is an SCM $\mathbb{C} = \langle \mathbf{X}, \mathbf{E}, \mathbf{N}, N_p \rangle$ where

$$\mathbf{X} = \{X_i \in \mathbb{B}^{|Q|}, Y_i \in \mathbb{B}^{|S|}\}_{i=0}^{T} \tag{6}$$

$$\mathbf{N} = \{N_{X_i} \in \mathbb{B}^{|Q| \times |Q|}, N_{Y_i} \in \mathbb{B}^{|Q| \times |S|}\}_{i=1}^{T} \cup N_{X_0} \in \mathbb{B}^{|Q|}. \tag{7}$$

and $X_t[i] = 1, i \in Q$ corresponds to the HMM being in the state i at time t and $Y_t[j] = 1, j \in S$ corresponds to the HMM emitting observation j at time t.



(a) A simple Hidden Markov Model $\mathbb{H}'$

(b) $T-$encoded SCM for $\mathbb{H}'$

Figure 2: A simple HMM and its SCM.

N_p is such that N_{X_t}, N_{Y_t} are the *IDD encoding of* $\mathbb{H}$ such that N_{X_t} is a binary matrix where $N_{X_t}[i,j] = 1$ with probability $P(i,j)$ and $\forall i, N_{X_t}[i,j] = 1$ for exactly one j. Similarly, $N_{X_0}[j] = 1$ with probability $\theta(j)$ and exactly one entry equaling 1. $N_{Y_t}[i,j] = 1$ with probability $O(i,j)$ and $\forall i, N_{Y_t}[i,j] = 1$ for exactly one j.

$$\mathbf{E} = \{X_0 := N_{X_0} \tag{8}$$

$$X_{t+1} := X_t \cdot N_{X_t} \tag{9}$$

$$Y_t := X_t \cdot N_{Y_t}\} \tag{10}$$

4.3 Total Causal Effect in the T-encoded SCM

We are ready to consider the total causal effect in $IU(\mathbb{H})$ for an HMM $\mathbb{H}$ and its relationship to p-causes. As shown in Figure 2b, the structure of $IU(\mathbb{H})$ for any $\mathbb{H}$ is the same. Therefore, when computing total causal effect on the same time steps, we can choose the same adjustment set.

Consider the HMM $\mathbb{H}$ in Figure 1 and the language for observing H on the third step: $\mathscr{L} = \bigcirc\bigcirc H$.

We know from Example 1, that MM is in the p-cause for $\mathscr{L}$. How might we check for a total causal effect for this prefix and language? In $\mathbb{C} = IU(\mathbb{H})$, the vertices $C = \{X_0, Y_0, X_1, Y_1\}$ represent the probability distribution of the states and observation in the first two time steps of $\mathbb{H}$. Since the MM is a prefix of length 2, its probability is in the probability distribution of C. Similarly, $E = \{X_2, Y_2\}$ represents the probability distribution of the states and observations at the 3rd time step and thus the probability of emitting $\Sigma^2 H$. Therefore, we will check for a total causal effect from C to E in $\mathbb{C}$.

To calculate a total causal effect, we let the adjustment set $Z = \emptyset$ because there are no backdoor paths from C to E; therefore, $P^{\mathbb{C};do(C:=c)}(E) = P^{\mathbb{C}}(E|C := c)$ (see [13] Proposition 6.41). We see in $\mathbb{H}$ for $E = \{S_3, H\}$ that $P_{\mathbb{H}}(S_1, M, S_2, M, S_3, H | S_1, M, S_2, M) \neq P_{\mathbb{H}}((S, \Sigma)^2, S_3, H)$. Since $P_{\mathbb{C}}$ entails $P_{\mathbb{H}}$, $P^{\mathbb{C}}(E := \{S3, H\} | C := \{S_1, M, S_2, M\}) \neq P^{\mathbb{C}}(E := \{S3, H\})$. Therefore, by Proposition 4, there is a total causal effect from C to E.

This does not let us conclude that p-causes are necessary for total causal effects. If we wanted to check for a total causal effect of a *non p-critical prefix*, say LL, we would select the same sets of vertices C, E, and we would find that there is also a total causal effect. In contrast, in the HMM $\mathbb{H}'$ in Figure 2a, with the same language $\mathscr{L}$, MM is also a p-critical prefix. Selecting the same E, C, we find that there is no total causal effect from C to E because $P^{\mathbb{C}}(E|C := c) = P^{\mathbb{C}}(E)$ for all choices of c.

We see that there is a total causal effect when C and E are dependent in $\mathbb{C}$, but not on a particular p-cause. We observe that $C = \{X_i, Y_i\}_{i=0}^{l}$ and $E = \{X_i, Y_i\}_{j=n}^{m}$ are independent when all paths in $\mathbb{H}$ from states reachable at time step l to states reachable at time step n are deterministic.

In summary, we find that that given this encoding, p-critical prefix $\not\Rightarrow$ total causal effect and total causal effect $\not\Rightarrow$ p-critical prefix.

References

[1] : .

[2] *Causal inference with multiple time series: principles and problems*, doi:10.1098/rsta.2011.0613. Available at `https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2011.0613`.

[3] Christel Baier, Clemens Dubslaff, Florian Funke, Simon Jantsch, Rupak Majumdar, Jakob Piribauer & Robin Ziemek (2021): *From Verification to Causality-based Explications*. Available at `http://arxiv.org/abs/2105.09533`. Number: arXiv:2105.09533 arXiv:2105.09533 [cs].

[4] Christel Baier, Florian Funke, Simon Jantsch, Jakob Piribauer & Robin Ziemek (2021): *Probabilistic Causes in Markov Chains*. In Zhe Hou & Vijay Ganesh, editors: *Automated Technology for Verification and Analysis*, Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 205–221, doi:10.1007/978-3-030-88885-5_14.

[5] Sander Beckers, Hana Chockler & Joseph Y. Halpern (2022): *A Causal Analysis of Harm*, doi:10.48550/ARXIV.2210.05327. Available at `https://arxiv.org/abs/2210.05327`.

[6] David Hume (1739): *A Treatise of Human Nature*. Clarendon Press.

[7] Amjad Ibrahim, Tobias Klesel, Ehsan Zibaei, Severin Kacianka & Alexander Pretschner (2020): *Actual Causality Canvas: A General Framework for Explanation-Based Socio-Technical Constructs*. doi:10.3233/FAIA200472.

[8] Thomas F Icard: *From Programs to Causal Models*, p. 10.

[9] Samantha Kleinberg: *A Logic for Causal Inference in Time Series with Discrete and Continuous Variables*.

[10] Samantha Kleinberg & George Hripcsak (2011): *A review of causal inference for biomedical informatics*. *Journal of biomedical informatics* 44, pp. 1102–12, doi:10.1016/j.jbi.2011.07.001.

[11] Judea Pearl (2009): *Causal inference in statistics: An overview*. *Statistics Surveys* 3(none), doi:10.1214/09-SS057. Available at `https://projecteuclid.org/journals/statistics-surveys/volume-3/issue-none/Causal-inference-in-statistics-An-overview/10.1214/09-SS057.full`.

[12] Judea Pearl & Dana Mackenzie (2018): *The Book of Why*. Basic Books, New York.

[13] Jonas Peters, Dominik Janzing & Bernhard Schölkopf (2017): *Elements of causal inference: foundations and learning algorithms*. Adaptive computation and machine learning series, The MIT Press, Cambridge, Massachuestts.

[14] Dawei Shi, Ziyang Guo, Karl Henrik Johansson & Ling Shi (2018): *Causality Countermeasures for Anomaly Detection in Cyber-Physical Systems*. *IEEE Transactions on Automatic Control* 63(2), pp. 386–401, doi:10.1109/TAC.2017.2714646. Conference Name: IEEE Transactions on Automatic Control.

[15] Zenna Tavares, James Koppel, Xin Zhang, Ria Das & Armando Solar-Lezama (2021): *A Language for Counterfactual Generative Models*. In: *Proceedings of the 38th International Conference on Machine Learning*, PMLR, pp. 10173–10182. Available at `https://proceedings.mlr.press/v139/tavares21a.html`. ISSN: 2640-3498.