Temporal Causality in Reactive Systems[*]

Norine Coenen[1], Bernd Finkbeiner[1], Hadar Frenkel[1],
Christopher Hahn[2], Niklas Metzger[1], and Julian Siber[1](✉)

[1] CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
{norine.coenen,finkbeiner,hadar.frenkel,
niklas.metzger,julian.siber}@cispa.de
[2] Stanford University, Stanford, USA
hahn@cs.stanford.edu

Abstract. Counterfactual reasoning is an approach to infer what causes an observed effect by analyzing the hypothetical scenarios where a suspected cause is not present. The seminal works of Halpern and Pearl have provided a workable definition of counterfactual causality for finite settings. In this paper, we propose an approach to check causality that is tailored to reactive systems, i.e., systems that interact with their environment over a possibly infinite duration. We define causes and effects as trace properties which characterize the input and observed output behavior, respectively. We then instantiate our definitions for ω-regular properties and give automata-based constructions for our approach. Checking that an ω-regular property qualifies as a cause can then be encoded as a hyperproperty model-checking problem.

1 Introduction

Causality plays an increasingly important role in computer science, e.g., to explain the behavior of a system [3,4,9,7], to establish accountability in multiagent systems [10], or to solve challenging algorithmic problems [21,2]. These approaches commonly draw upon the rich philosophical literature that has laid the foundation for *counterfactual reasoning* [23,19], a method of establishing causal relationships between events. According to this line of reasoning, a cause is an event such that, if it had not happened, the effect would not have happened either. A rigorous formalization of counterfactual causality has been proposed by Halpern and Pearl [16]. This formalization is first and foremost concerned with models that can be described by a finite set of variables. When naively applying it to reactive systems that interact with their environment continuously, however, the analysis may infer that an infinite number of events (variable valuation at time step) are causes for an observed effect, falling short of providing the intended comprehensible explanation [17].

In this paper, we therefore propose an approach to causal analysis in reactive systems that provides a symbolic description of causes. We define counterfactual

[*] This work was funded by DFG grant 389792660 as part of TRR 248 – CPEC and by the German Israeli Foundation (GIF) Grant No. I-1513-407./2019.

causality on the basis of trace properties (Section 4), i.e., causes are properties of a given input sequence, and effects are properties of the corresponding output sequence, and apply this definition to ω-regular properties to obtain concrete automata-based constructions (Section 5). As one of our building blocks, we adapt counterfactual automata [9] so they generate all relevant counterfactual traces in our setting. Our definitions are sufficiently general to be instantiated by a variety of temporal logics, such as LTL [25] or QPTL [27]. This general approach allows us to leverage the significant previous work on temporal logics and for the usual trade-off between expressiveness and decidability.

Our notion of causality is an *actual* kind of causality in the spirit of Halpern and Pearl [16]. This means we provide a precise description of the temporal behavior responsible for the effect on a given, *actual trace* of the reactive system. This actual trace can, for example, be provided as a counterexample by a model checker, where the effect then is the violation of the specification. We define what it means to *intervene* on the cause property of an actual trace, i.e., how to modify the trace such that the property is not satisfied anymore, but the resulting counterfactual trace is still sufficiently close to the actual trace to comply with the closest possible worlds principle [23]. We then further allow for *contingencies* as introduced by Halpern and Pearl [16], to isolate the exact causal behavior in case of preemption of other potential causes.

Previous approaches to provide symbolic descriptions of counterfactual causes use an event-based logic [22,6], which allows reasoning about the order of events, but cannot, e.g., specify at which time step a causal input occurs. In contrast, our framework is only limited by the expressiveness of the logic used to describe the causal trace properties. We study a decidable instantiation of our definitions with Quantified Propositional Temporal Logic (QPTL), an extension of LTL with quantified atomic propositions. Causes can be identified as a temporal property (see Section 3 for an example). Moreover, the event-based approaches are restricted to finitely observable effects [22] or define a system-level causality that does not consider the causal dependencies on a given, actual trace [6]. In comparison, our approach allows for a significantly more precise description of the temporal causal behavior on an observed system trace.

As an intriguing theoretical result, we show that when a candidate cause for an effect is given as a trace property, checking whether it is indeed the actual cause on a trace of a system cannot be stated as a trace property, which formalizes previous observations on counterfactual causality [10]. The result motivates us to consider causality as a hyperproperty [8] in our approach. In particular, we show that verifying ω-regular causality on lasso-shaped traces is decidable via HyperQPTL model checking.

2 Preliminaries

Systems and Traces. We model a reactive system as a (nondeterministic) *Moore machine* [24] $\mathcal{T} = (S, s_0, AP, \delta, l)$ where S is a finite set of states, $s_0 \in S$ is the initial state, $AP = I \mathbin{\dot\cup} O$ is the set of atomic propositions consisting of inputs I and outputs O, $\delta : S \times 2^I \to 2^S$ is the transition function determining a set of successor states for a given state and input, and $l : S \to 2^O$ is the labeling function mapping each state to a set of outputs. A path $s = s_0 s_1 \ldots \in S^\omega$ of $\mathcal{T}$ is an infinite sequence with $s_{i+1} \in \delta(s_i, I_i)$ for all $i \in \mathbb{N}$ and for some $I_i \subseteq I$, we assume there exists such $s' \in \delta(s, Y)$ for all $s \in S$ and $Y \subseteq I$. The corresponding trace is $\pi = \pi_0 \pi_1 \pi_2 \ldots \in (2^{AP})^\omega$, such that $\pi_i = I_i \cup l(s_i)$ for the I_i used by δ. With *traces*$(\mathcal{T})$, we denote the set of all traces of $\mathcal{T}$. For two subsets of atomic propositions V, $W \subseteq AP$, let $V|_W = V \cap W$ and $\pi|_W = \pi_0|_W \, \pi_1|_W \ldots$ for some trace π. We say a trace π is *lasso-shaped*, if there exist $i, j = i+1, k \in \mathbb{N}$ such that $\pi = \pi_0 \ldots \pi_i \cdot (\pi_j \ldots \pi_k)^\omega$. For some subset $A \subseteq AP$, we call a set of traces $\mathsf{P} \subset (2^A)^\omega$ a *trace property*. A trace π satisfies P, denoted by $\pi \vDash \mathsf{P}$ iff $\pi|_A \in \mathsf{P}$.

QPTL and HyperQPTL. HyperQPTL [26] is a temporal logic that can express ω-regular hyperproperties. HyperQPTL is derived from linear-time temporal logic (LTL) [25] by adding explicit quantification over atomic propositions (leading to quantified propositional temporal logic (QPTL) [27]) and explicit quantification over trace variables (for relating multiple traces):

$$\varphi ::= \forall \pi.\, \varphi \mid \exists \pi.\, \varphi \mid \forall q.\, \varphi \mid \exists q.\, \varphi \mid \ \psi$$
$$\psi ::= a_\pi \mid q \mid \neg \psi \mid \psi \wedge \psi \mid \bigcirc \psi \mid \psi \,\mathcal{U}\, \psi$$

for a trace variable $\pi \in \mathcal{V}$, fresh atomic proposition $q \notin \mathrm{AP}$, and atomic proposition $a \in \mathrm{AP}$. We also consider the usual derived Boolean ($\vee$, $\rightarrow$, $\leftrightarrow$) and temporal operators ($\varphi \,\mathcal{R}\, \psi \equiv \neg(\neg\varphi \,\mathcal{U}\, \neg\psi)$, $\Diamond \varphi \equiv \mathit{true} \,\mathcal{U}\, \varphi$, $\Box \varphi \equiv \mathit{false} \,\mathcal{R}\, \varphi$). The semantics of HyperQPTL is defined with respect to a time point i, a set of traces Tr and a trace assignment $\Pi : \mathcal{V} \to Tr$ that maps trace variables to traces. To update the trace assignment so that it maps trace variable π to trace t, we write $\Pi[\pi \mapsto t]$. HyperQPTL introduces an auxiliary trace variable π_q for every quantified atomic proposition q. The semantics is as follows:

$$\begin{array}{lll}
\Pi, i \vDash_{Tr} a_\pi & \text{iff} & a \in \Pi(\pi)[i] \\
\Pi, i \vDash_{Tr} q & \text{iff} & q \in \Pi(\pi_q)[i] \\
\Pi, i \vDash_{Tr} \neg\varphi & \text{iff} & \Pi, i \nvDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \varphi \wedge \psi & \text{iff} & \Pi, i \vDash_{Tr} \varphi \text{ and } \Pi, i \vDash_{Tr} \psi \\
\Pi, i \vDash_{Tr} \bigcirc \varphi & \text{iff} & \Pi, i+1 \vDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \varphi \,\mathcal{U}\, \psi & \text{iff} & \exists j \geq i.\, \Pi, j \vDash_{Tr} \psi \wedge \forall i \leq k < j.\, \Pi, k \vDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \forall \pi.\, \varphi & \text{iff} & \text{for all } t \in Tr \text{ it holds that } \Pi[\pi \mapsto t], i \vDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \exists \pi.\, \varphi & \text{iff} & \text{there is some } t \in Tr \text{ such that } \Pi[\pi \mapsto t], i \vDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \forall q.\, \varphi & \text{iff} & \text{for all } t \in (2^{\{q\}})^\omega \text{ it holds that } \Pi[\pi_q \mapsto t], i \vDash_{Tr} \varphi \\
\Pi, i \vDash_{Tr} \exists q.\, \varphi & \text{iff} & \text{there is some } t \in (2^{\{q\}})^\omega \text{ it holds that } \Pi[\pi_q \mapsto t], i \vDash_{Tr} \varphi \ .
\end{array}$$

The semantics of a QPTL formula φ can be derived from HyperQPTL formula $\forall \pi.\ \varphi_\pi$, where φ_π is obtained by indexing all atomic propositions in φ with π.

Actual Causality. We shortly outline actual causality originally proposed by Halpern and Pearl [16], in the version modified by Halpern [15]. A *causal model*

$\mathcal{M} = (\mathcal{S}, \mathcal{F})$ is defined by a *signature* $\mathcal{S}$ and set of *structural equations* $\mathcal{F}$. A signature $\mathcal{S}$ is a tuple $(\mathcal{E}, \mathcal{V}, \mathcal{R})$, where $\mathcal{E}$ is a set of *exogenous* variables, $\mathcal{V}$ is a set of *endogenous* variables, and $\mathcal{R}$ defines the *range* of possible values $\mathcal{R}(Y)$ for all variables $Y \in \mathcal{E} \cup \mathcal{V}$. For some context $\vec{u}$, the value of an exogenous variable is determined by factors outside of the model, while the value of some endogenous variable X is defined by the associated structural equation $f_X \in \mathcal{F}$.

Definition 1. *$\vec{X} = \vec{x}$ is an* actual cause *of φ in $(\mathcal{M}, \vec{u})$, if the following holds.*

AC1: $(\mathcal{M}, \vec{u}) \vDash \vec{X} = \vec{x}$ and $(\mathcal{M}, \vec{u}) \vDash \varphi$, i.e., both cause and effect are true in the actual world, and

AC2: There is a set $\vec{W}$ of variables in $\mathcal{V}$ and a setting $\vec{x}'$ of the variables in $\vec{X}$ such that if $(\mathcal{M}, \vec{u}) \vDash \vec{W} = \vec{w}$, then $(\mathcal{M}, \vec{u}) \vDash [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\varphi$, and

AC3: $\vec{X}$ is minimal, i.e. no subset of $\vec{X}$ satisfies AC1 and AC2.

Intuitively, AC2 means that after *intervening* on the actual world such that the cause $\vec{X} = \vec{x}$ is not satisfied, the effect is not satisfied either. AC2 allows further modification through the notion of *contingencies.* The contingency $\vec{W}$ can, in the hypothetical world, be reset to the original value it takes in the actual world, even when the intervention on $\vec{X}$ may have altered it.

3 Motivating Example

As an illustration of our approach, we consider the problem of identifying a spurious arbiter. The purpose of an arbiter is to organize mutually exclusive access to a shared resource by eventually answering a request of this resource with a grant. This may be achieved by simply giving grants in a round-robin strategy, regardless of incoming requests. Such spurious and inefficient behavior is unwanted in practice but may result from a sub-optimal specification as input to a reactive synthesis procedure. Our causality-checking approach can identify it by checking whether, e.g., a request r_1 is a cause for a grant g_1 by checking whether the temporal property $\Diamond r_1$ *causes* the observed behavior described by the temporal property $\Diamond g_1$ on a given trace π.

The causal analysis utilizes counterfactual reasoning: if on the traces π' of the system that are similar to π, but where the cause-property $\Diamond r_1$ is not satisfied, the effect-property $\Diamond g_1$ also does not occur, we can infer a causal relationship between the two properties on input and output sequence. As an example, consider the following trace of the system depicted on the left in Figure 1: $\pi_1 = (\{r_1, g_1\}\{r_0, g_0\})^\omega$. Counterfactual reasoning now requires us to consider similar traces where no r_1 occurs, i.e., the negation of the cause property, which is $\Box\neg r_1$, holds. In particular, since we consider sequences that are still sufficiently similar to π, we require that the sequence does not change the occurrences of r_0. Consequently, the counterfactual trace we are interested in is given by $\pi_1' = (\{g_1\}\{r_0, g_0\})^\omega$.

As we can see, the effect still occurs on π_1', therefore $\Diamond r_1$ is *not* a cause for $\Diamond g_1$ on π_1 in the spurious arbiter. In contrast, consider the arbiter depicted

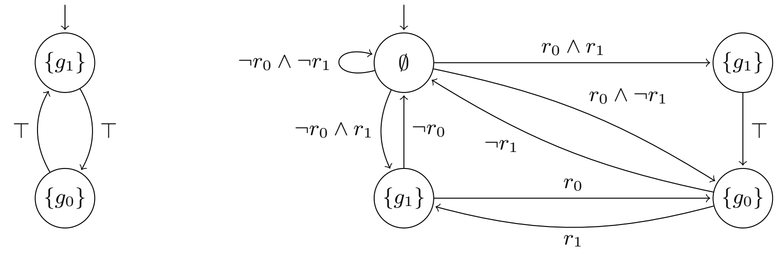

Fig. 1: The models of a spurious (left) and non-spurious arbiter (right). Here, edges are labeled with symbolic constraints, e.g., $\neg r_1$ for all sets without r_1.

on the right in Figure 1, which works as expected, i.e., only gives out grants upon receiving a request. Let us check whether the causal relationship between the properties from above holds on the trace $\pi_2 = \{r_1\} \cdot (\{r_0, g_1\}\{r_1, g_0\})^\omega$. Here, applying the same counterfactual reasoning technique from before actually yields the following trace, which does not satisfy the effect property: $\pi_2' = \{\} \cdot (\{r_0\}\{g_0\})^\omega$. Hence, we can infer that $\Diamond r_1$ is a cause for $\Diamond g_1$ on π_2.

Even in correct systems, our causal analysis allows further insight into which exact behavior is responsible for a certain observed effect. In particular, *contingencies* allow to isolate a cause in the presence of multiple potential causes on the trace that were preempted. To illustrate, consider again the right arbiter from Figure 1 and the following trace: $\pi_3 = \{r_1\}\{r_1, g_1\} \cdot \{\}^\omega$. We may now ask whether it is the first or the second r_1 that causes the effect $\Diamond g_1$ on π_3. When considering only the naive counterfactual trace $\pi_3' = \{\}\{r_1\}\{g_1\} \cdot \{\}^\omega$ during analysis of formula r_1 a problem occurs. In π_3', the second request takes effect even though it had no effect in π_3. Contingencies now allow us to reset certain parts of the counterfactual trace back to the actual trace. In particular, we are allowed to change the state and outputs at the third position to their value in π_3, which yields the following counterfactual trace under the contingency: $\pi_3'' = \{\}\{r_1\} \cdot \{\}^\omega$. Since π_3'' does not satisfy the effect, the analysis establishes a causal relationship between the property r_1 of the input sequence and the property $\Diamond g_1$ of the output sequence. Note that considering the alternative, intuitively more precise effect-property $\bigcirc g_1$ leads to the same result without the need for contingencies. Hence, contingencies allow us to precisely infer the causal behavior even if the effect is described in a more general manner.

4 Property Causality

In this section we lift the definitions of Halpern and Pearl to the setting of causes and effects given as general trace properties. We define when some temporal behavior on the input sequence of a reactive system is considered a cause for some temporal behavior observed on the output sequence. We assume a cause $\mathsf{C} \subseteq (2^I)^\omega$ to be a trace property reasoning only over the input variables of the system and an effect $\mathsf{E} \subseteq (2^O)^\omega$ to be a trace property ranging over the output variables. We call such properties *cause property* and *effect property*, respectively. In an abuse of notation, we will sometimes use QPTL formulas for C and E in this section when we interpret their language as a trace property.

In order to lift Definition 1 to the setting of both infinite traces and infinite sets of traces for cause and effect, we need to be able to reason about *interventions* (Section 4.1), i.e., how to modify the actual trace such that the cause property does not hold anymore; and *contingencies* (Section 4.2), that allow to infer the exact causal behavior when it preempts other potential causes. We then can introduce our full definition for temporal causality in Section 4.3.

4.1 Interventions

Recall that at the core of counterfactual reasoning lies the idea that if the cause had not appeared on the given trace, then the effect would not have happened either. Hence, as a first step we need to define how the counterfactual traces, i.e., the traces that are just like our given trace, but where the cause-property C is not satisfied, look like. We follow the classic theory of closest possible worlds introduced by Lewis [23] to characterize a set of counterfactual traces that lie just outside of C. For that, we are interested in defining the minimal sets that modify the given trace such that some cause-property C is not satisfied anymore. We call such a set an *intervention*. Following Halpern and Pearl definition, a set is minimal if none of its subsets alone suffice to change the evaluation of C on π. However, for the case of general trace properties as cause and effect, this notion would not allow us to find any minimal interventions.

Example 2. Consider again the non-spurious arbiter from Figure 1 (right), the cause $\mathsf{C} = \square\Diamond r_1$ and the effect $\mathsf{E} = \square\Diamond g_1$, and the trace $\pi = \{r_1\} \cdot \{r_1, g_1\}^\omega$. Traces that falsify the effect are traces with only finitely many occurrences of r_1. However, if we follow the subset definition for minimal interventions (see Definition 1), and values of atomic propositions at time point as the respective variables, we get that each trace of the form $\{r_1\}\{r_1, g_1\}^k\{g_1\} \cdot \{\}^\omega$ has a trace with less changes with respect to π, that also falsify the effect, e.g., $\{r_1\}\{r_1, g_1\}^{k+1}\{g_1\} \cdot \{\}^\omega$. Therefore, if we look for minimal interventions using this naive reasoning, we will never find counterfactual traces.

As a solution, we link the satisfaction of the cause property to a distance measure that partially orders counterfactual traces with respect to π. Because this concept is applicable beyond the models and logics considered in this paper, we give a general definition that can be applied to other domains as well.

Formally, we require the existence of a distance measure $<^{\mathsf{C}}_{\pi}$ that conforms with the underlying logic to detect minimal intervention traces. Such traces σ are the closest to π according to $<^{\mathsf{C}}_{\pi}$ that do not satisfy C, i.e., there is no $\rho \notin \mathsf{C}$ such that $\rho <^{\mathsf{C}}_{\pi} \sigma$. Generally, multiple traces might satisfy this criterion, so we define a set of minimal interventions.

Definition 3 (Intervention Set). *Let* C *be a cause property, let* $\pi \vDash \mathsf{C}$ *be a trace, and let* $<^{\mathsf{C}}_{\pi}$ *be a distance measure that partially orders traces with respect to* π*. The* set V^{C}_{π} of interventions *on* C *with respect to* π *contains exactly all minimal interventions with respect to* π *according to* $<^{\pi}_{\mathsf{C}}$*. That is*

$$V^{\mathsf{C}}_{\pi} = \{\sigma \notin \mathsf{C} \mid \forall \rho \notin \mathsf{C}.\ \rho \not<^{\mathsf{C}}_{\pi} \sigma\} \enspace .$$

4.2 Contingencies

Next, we discuss the treatment of *contingencies* in reactive systems. The motivation behind contingencies is to isolate the truly causal behavior when there is preemption of other potential causes on the actual trace. Contingencies allow certain variables in the counterfactual trace to be reset to their value in the actual trace, in this way mimicking the fact that the second potential cause was preempted in the actual trace. To fully account for this preemption, it is not sufficient that only the output value at a single position is changed to the value in the actual trace: the future dynamics have to respect the contingency by additionally *changing the state* the trace is in when a contingency is evoked.

For a given counterfactual trace, we inductively define the resulting contingencies. Here, we assume a transition relation for the system that is not necessarily memoryless, as we consider general trace logics for now. However, we do assume that transitions only depend on the history of the trace and not on its future. This corresponds to the non-recursive models assumed by Halpern and Pearl. We thus extend the definition of transition function for Moore machines, given in Section 2, to a transition function that relies on the whole sequence of inputs and outputs observed so far.

Definition 4 (Contingency Set). *Let* $\delta^* : (2^I \times 2^O)^* \times 2^I \to 2^O$ *be a function that returns the possible next outputs* (2^O) *according to the history of the trace and the current input* (2^I)*, modeling the behavior of the system. Given an intervention trace* σ *and an original trace* π*, we define the* contingency set C^{σ}_{π} *where* $\pi' = \pi'_0\pi'_1 \ldots \in (2^I \times 2^O)^\omega$ *is in* C^{σ}_{π} *if the following two conditions hold:*

1. $\forall j \in \mathbb{N} :\ \ \pi'_j \cap 2^I = \sigma_j \cap 2^I$*; That is,* π' *has the same input sequence as* σ*.*
2. $\forall j \in \mathbb{N} :\ \ (o \in \pi'_j \leftrightarrow o \in \delta^*(\pi'_0 \cdots \pi'_{j-1} \cdot (\pi'_j \cap 2^I))) \vee (o \in \pi_j)$*; That is, the output sequence of* π' *is determined according to the behavior of the system, together with"jumps" to the original trace* π*. Note that since the input sequence of* σ *and* π' *is the same, it holds that until the first jump to* π*, the output sequence of* σ *and* π' *is also the same.*

Since a contingency only allows to reset outputs to their value in the actual trace, the set of traces under a contingency is defined relative to the actual trace π. The trace under the counterfactual input sequence σ, without modifications, is always part of the contingency set. Starting from this trace, contingencies can be enforced at infinitely many positions.

4.3 Actual Causality for Trace Properties

Minimality of the cause is defined simply based on strict set inclusion, and provides the last condition for our following definition of property-based causality.

Definition 5 (Property Causality). *Let $\mathcal{T}$ be a system, $\pi \in traces(\mathcal{T})$ a trace, $\mathsf{C} \subseteq (2^I)^\omega$ a cause property, and $\mathsf{E} \subseteq (2^O)^\omega$ an effect property. We say that C is a* cause *of E on π in $\mathcal{T}$ if the following three conditions hold:*

PC1: $\pi \models \mathsf{C}$ *and* $\pi \models \mathsf{E}$*, i.e., cause property and effect property are satisfied by the actual trace.*

PC2: *For every counterfactual input sequence $\sigma \in V_\pi^{\mathsf{C}}$, there is some contingency $\pi' \in C_\pi^\sigma$ s.t. $\pi' \not\models \mathsf{E}$, i.e., the counterfactual trace under contingency does not satisfy the effect property.*

PC3: *There is no C' s.t. $\mathsf{C}' \subset \mathsf{C}$ and C' satisfies PC1 and PC2.*

As a consequence of our treatment of minimality, there is always a maximal cause-property $\mathsf{C}_{max} = (2^I)^\omega$ that trivially satisfies PC1 and PC2. On the other hand, the minimal relevant cause property for a given trace π is $\mathsf{C}_{min} = \{\pi|_I\}$, i.e., the input sequence of the trace itself. This is because the empty set will never qualify for PC1. However, this does not imply that there is a well-defined minimal cause in all cases, because if the considered properties are expressive enough, it may be possible to find a subset that satisfies PC1 and PC2 for any candidate cause property, thus falsifying PC3.

It has been conjectured before that finding causes cannot be stated as a trace property [10]. This hypothesis has intuitive appeal because most notions of causality relate the actual world with counterfactual worlds based on certain similarity metrics. For our proposed notion of trace-based causality, we answer this intriguing question affirmatively in the following theorem and show that even deciding whether a cause candidate is an actual cause cannot be stated as a trace property.

Theorem 6. *Given a cause-property C, an effect-property E, and some trace π, there is no trace-property P such that for all systems $\mathcal{T}$ with $\pi \in traces(\mathcal{T})$ it holds that $\mathcal{T} \models \mathsf{P}$ iff C is a cause for E on π in $\mathcal{T}$.*

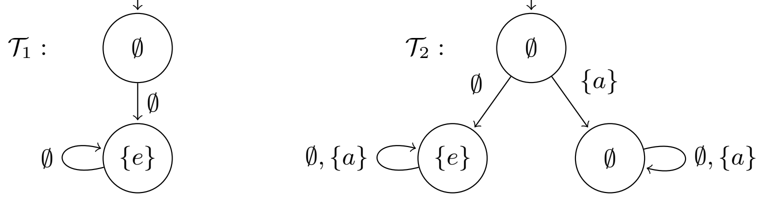

Fig. 2: The systems $\mathcal{T}_1$ and $\mathcal{T}_2$ used in the proof of Theorem 6.

Proof. By contradiction. Assume there is such a trace-property P for the cause-property $\mathsf{C} = \neg a$, the effect-property $\mathsf{E} = \Diamond e$, and the trace $\pi = \{\}\{e\}^\omega$. Now, consider the two systems depicted in Figure 2: $\mathcal{T}_1$ with $I_1 = \{\}$ and $\mathcal{T}_2$ with $I_2 = \{a\}$, and $O_1 = O_2 = \{e\}$. We have that C is a cause for E on π in $\mathcal{T}_2$, because we can avoid E in this system with the counterfactual input sequence $\sigma = \{a\}\{\}^\omega \in V_\pi^{\mathsf{C}}$. Note that contingencies do not matter in both systems because they can only set the trace to a state which immediately satisfies E. In $\mathcal{T}_1$, however, E cannot be avoided at all, hence C is not a cause for E in $\mathcal{T}_1$. However, since $\mathit{traces}(\mathcal{T}_1) \subset \mathit{traces}(\mathcal{T}_2)$, we have that $\mathcal{T}_2 \vDash \mathsf{P}$ implies $\mathcal{T}_1 \vDash \mathsf{P}$. It follows that C has to be a cause in $\mathcal{T}_1$, which contradicts the assumption. □

In this section, we have presented a general framework that establishes causal relationships between temporal properties given as sets of traces, on a given actual trace. The key idea is to link satisfaction of the property to a distance measure over potential counterfactual traces to obtain meaningful interventions, and to allow for contingencies based on relaxing the dynamics of the model such that it can jump back to states of the actual trace. The proposed concept can conceivably be applied to a variety of models and corresponding logics with a linear-time semantics. However, to allow algorithmic reasoning about the proposed property causality, it is of course necessary to fix a finite representation of the infinite traces and infinite sets, as we do in the following section.

5 Checking ω-Regular Causality

In this section we provide a decision procedure that allows us to check ω-regular causes with respect to ω-regular effects, i.e., verify whether a given candidate cause property is indeed a cause for an observed effect property on an actual trace. We use causes and effects given in the logic QPTL (see Section 2), which is equivalent to the class of ω-regular properties. Note that Linear Temporal Logic (LTL), which is one of the standard specification languages for specifying temporal properties in reactive systems, is subsumed by QPTL. We further assume that our actual trace π is given in a finite, lasso-shaped representation (as defined in Section 2). This is a common assumption when verifying LTL properties, since if there exists a violation, in particular there exists also a lasso-shaped violation. Model-checking tools (e.g. [18]) usually return such a structured trace. Due to space constraints, we omit language-theoretic definitions in this section and provide definitions directly as HyperQPTL properties, as this allows us to directly reason about their decidability.

5.1 Interventions

We now formalize our discussion of interventions from Section 4.1 for QPTL. Our distance measure closely mirrors the original minimality criterion of Halpern and Pearl over sets of variables (see Definition 1), i.e., a trace ρ is closer to the actual trace π than some other trace σ if the events differing between π and ρ are a strict subset of the events differing between π and σ. We can formalize this with the following HyperQPTL property.

$$\psi_{min}(\pi, \rho, \sigma) = \Big(\Box \bigwedge_{a \in I} \big((a_\rho \not\leftrightarrow a_\pi) \rightarrow (a_\sigma \not\leftrightarrow a_\pi)\big)\Big) \wedge \Big(\Diamond \bigvee_{a \in I} (a_\rho \not\leftrightarrow a_\sigma)\Big)$$

However, to avoid the issue discussed in Example 2, we only order counterfactual traces that share the same *rejection structure* with respect to the cause-property C, i.e., if they satisfy the right-hand subformulas of every $\mathcal{U}$ (and the derived temporal operators $\Diamond$ and $\Box$) appearing in $\neg\mathsf{C}$ at the same positions.[3] To formalize this requirement as a HyperQPTL property, let $\varphi^{\mathcal{U}_1}_{\neg\mathsf{C}}(\pi), \ldots, \varphi^{\mathcal{U}_n}_{\neg\mathsf{C}}(\pi)$ be these subformulas appearing in $\neg\mathsf{C}$, with their atomic propositions indexed by the parameterized π. The two traces σ and ρ have the same rejection structure with respect to C if they satisfy the following HyperQPTL property $\psi^{\mathsf{C}}_{struct}(\rho, \sigma)$.

$$\psi^{\mathsf{C}}_{struct}(\rho, \sigma) = \bigwedge_{i \in [1,n]} \Box \big(\varphi^{\mathcal{U}_i}_{\neg\mathsf{C}}(\rho) \leftrightarrow \varphi^{\mathcal{U}_i}_{\neg\mathsf{C}}(\sigma)\big)$$

Finally, we obtain an instantiation of the partial order $<^{\mathsf{C}}_{\pi}$ for QPTL such that for two traces σ, ρ: $\rho <^{\mathsf{C}}_{\pi} \sigma$ iff $\psi_{min}(\pi, \rho, \sigma) \wedge \psi^{\mathsf{C}}_{struct}(\rho, \sigma)$ holds. Note that since we only compare traces with the same rejection structure, we can always find minimal interventions, except if the cause property is a tautology.

Example 7. To illustrate how the above solves the problem raised in Example 2, consider the traces $\sigma = \{r_0, r_1\}^k \cdot \{r_0\}^\omega$ and $\rho = \{r_0, r_1\}^{k+1} \cdot \{r_0\}^\omega$, both in relation to $\pi = \{r_0, r_1\}^\omega$ and the cause-property $\mathsf{C} = \Box\Diamond r_1$. While we still have that $\psi_{min}(\pi, \rho, \sigma)$ holds, we have that $\psi^{\mathsf{C}}_{struct}(\rho, \sigma)$ does not hold because σ and ρ satisfy $\Box\neg r_1$ at different positions. Hence, σ, ρ are not ordered by $<^{\mathsf{C}}_{\pi}$ so both are in V^{C}_{π}. However, minimality still plays a key role such that we cannot manipulate r_0 in any valid intervention. Consider $\sigma' = \{r_1\}^k \cdot \{\}^\omega$. We have that $\psi^{\mathsf{C}}_{struct}(\sigma', \sigma)$ holds since both traces have the same rejection structure with respect to $\neg\mathsf{C}$. However, the changes in σ imply changes in σ', but not in the other direction. Hence, $\psi_{min}(\pi, \sigma, \sigma')$ and $\sigma <^{\mathsf{C}}_{\pi} \sigma'$, so only σ is in V^{C}_{π}.

5.2 Contingencies

We formalize the behavior of contingencies for ω-regular properties using a generalization of counterfactual automata as introduced by Coenen et al. [9]. In the original definition, they are restricted to systems whose states are uniquely labeled and which have a state for every output combination. We avoid this restriction by leveraging Halpern and Pearl's thoughts on models in which there exists no unique solution to the structural equations [16]. In these cases, they

[3] This is related to the notion of acceptance for words in nondeterministic Büchi automata [5], which recognize the class of ω-regular languages.

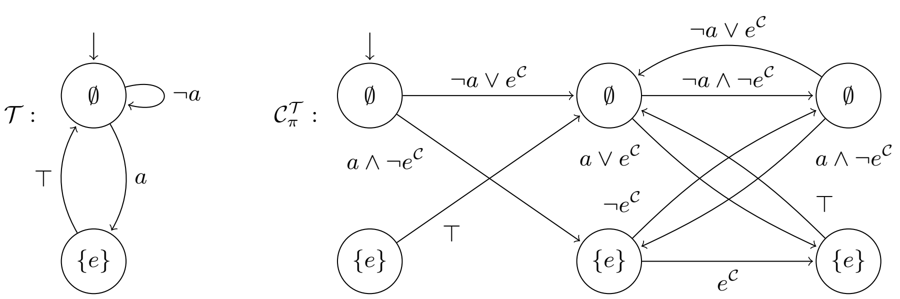
$\mathcal{T}$:
$\emptyset$
$\neg a$
$\top$
a
$\{e\}$
$\mathcal{C}_\pi^\mathcal{T}$:
$\neg a \vee e^\mathcal{C}$
$\neg a \wedge \neg e^\mathcal{C}$
$a \wedge \neg e^\mathcal{C}$
$a \vee e^\mathcal{C}$
$\neg e^\mathcal{C}$
$\top$
$e^\mathcal{C}$

Fig. 3: System $\mathcal{T}$ and the counterfactual automaton $\mathcal{C}_\pi^\mathcal{T}$ for $\pi = \{\} \cdot (\{a\}\{a,e\})^\omega$.

propose to use existential quantification over the solutions. In the same manner, we allow changing the underlying state of the trace to any state that is labeled with the right outputs. Since this means there might be several successor states for a given input and contingency combination, we formalize the counterfactual automaton as a nondeterministic Moore machine.

Definition 8 (Counterfactual Automaton [9] for General Systems). *Let $\mathcal{T} = (S, s_0, AP, \delta, l)$ be a system and $\pi = \pi_0 \dots \pi_i \cdot (\pi_j \dots \pi_k)^\omega \in traces(\mathcal{T})$ be a lasso-shaped trace. The* counterfactual automaton *for π and $\mathcal{T}$ is a Moore machine $\mathcal{C}_\pi^\mathcal{T} = (S^\mathcal{C}, s_0^\mathcal{C}, I^\mathcal{C} \cup O, \delta^\mathcal{C}, l^\mathcal{C})$, such that:*

- *$S^\mathcal{C} = S \times \{0 \dots k\}$, we have k copies of the original system;*
- *$s_0^\mathcal{C} = (s_0, 0)$, paths start in the initial state of the first copy;*
- *$I^\mathcal{C} = I \cupdot \{o^\mathcal{C} \mid o \in O\}$, additional inputs for setting an output as contingency;*
- *$(s', n') \in \delta^\mathcal{C}((s,n), Y)$ iff the following holds:*
  1. *if $n = k$ then $n' = j$ else $n' = n + 1$, and*
  2. *there is some $s'' \in \delta(s, Y|_I)$ such that for all $o^\mathcal{C} \in Y$: $o \in l(s') \leftrightarrow o \in \pi_n$ and for all $o^\mathcal{C} \notin Y$: $o \in l(s') \leftrightarrow o \in l(s'')$;*
- *$l^\mathcal{C}((s,k)) = l(s)$, the labeling function is based on the original states.*

The counterfactual automaton simulates arbitrary traces of the original system $\mathcal{T}$, which additionally can at every position choose to invoke a contingency through the additional inputs in $I^\mathcal{C}$ (see Condition 2), i.e., change the subsequent path to a state whose label is as of the next state determined by the original transition relation δ, but with all $o \in O$ that have their corresponding input $o^\mathcal{C} \in I^\mathcal{C}$ enabled set to their value as in π. Since π is of a finite, lasso-shaped form of length $k+1$, we can construct this behavior based on $k+1$ copies of the original system and enforce that a path proceeds from one copy to the next in every step (see Condition 1). In this way, the traces of the counterfactual automaton describe the set of all possible counterfactual traces under arbitrary contingencies. The idea is to then pick the subset of traces whose input behavior corresponds to interventions as defined in the previous section.

Example 9. To illustrate the idea of counterfactual automata, consider system $\mathcal{T}$ depicted in Figure 3 and the trace $\pi = \{\} \cdot (\{a\}\{a,e\})^\omega$. Since the trace has three positions, the counterfactual automaton $\mathcal{C}_\pi^\mathcal{T}$ consists of three copies of the original system. It has a single additional input $e^\mathcal{C}$. Every step in $\mathcal{C}_\pi^\mathcal{T}$ moves through the copies according to π's lasso structure, e.g., the copies of the prefix are only visited once on every path. If a trace does not set a contingency, it directly corresponds to a trace in $\mathcal{T}$, e.g., the trace $\pi' = \{\}^\omega$ is also a trace of $\mathcal{C}_\pi^\mathcal{T}$. However, setting contingencies allows to build traces in $\mathcal{C}_\pi^\mathcal{T}$ that do not have a corresponding trace in $\mathcal{T}$, e.g., $\pi'' = \{\} \cdot (\{e^\mathcal{C}\}\{e\})^\omega \notin \mathit{traces}(\mathcal{T})$.

Note that there might be several states that satisfy Condition 2 in Definition 8. This means that the precision of our causal analysis depends on how much state information the system exposes via its outputs: If every state is uniquely labeled, then a contingency can only set the trace to the state as in the actual trace and there is no ambiguity. Any system can be made amenable to this with auxiliary output variables for the state space.

5.3 Minimality

Our approach to check the minimality of a given ω-regular cause property is based on the observation that it suffices to find exactly one trace in C that can be characterized by an ω-regular language R, which can be removed from C to obtain a smaller ω-regular cause-property C'. This observation is formalized in the following lemma.

Lemma 10. *Let π be a trace and C a cause property that satisfies PC1 and PC2 for some effect-property E. Then, C satisfies PC3 if and only if $\forall \sigma \in \mathsf{C}, \forall \pi' \in C_\pi^\sigma.\ \pi' \vDash \mathsf{E}$ and $\forall \sigma \in \mathsf{C}, \forall \sigma' \in V_\pi^\mathsf{C}.\ \sigma' \not<_\pi^\mathsf{C} \sigma$.*

Proof. "$\Longrightarrow$": By contraposition. Let us distinguish two cases based on which of the conjuncts is false. We show that in both cases we can remove some ω-regular property R from C such that PC1 and PC2 still hold.

For the first case, assume there exist some $\sigma \in \mathsf{C}$ and $\pi' \in C_\pi^\sigma$ such that $\pi' \nvDash \mathsf{E}$. Since the quantified formula can be expressed as a HyperQPTL property of $\mathcal{C}_\pi^\mathcal{T}$, which encodes C_π^σ, we know that in particular there exists a witness σ that can be characterized by an ω-regular property $\mathsf{R} = \{\sigma\}$. We have $\sigma \neq \pi|_I$ because $C_\pi^\pi = \{\pi\}$ and $\pi \vDash \mathsf{E}$. Hence $\mathsf{C}' = \mathsf{C} \setminus \mathsf{R}$, which is again an ω-regular property, satisfies PC1, as $\pi \in \mathsf{C}'$. For PC2, consider the set $V' = \{\sigma' \in V_\pi^\mathsf{C} \mid \sigma <_\pi^\mathsf{C} \sigma'\}$ of intervention traces that follow the same structure as σ (and are less minimal). If V' is not empty, we have $V_\pi^{\mathsf{C}'} = (V_\pi^\mathsf{C} \setminus V') \cup \{\sigma\}$, as σ is now a more minimal intervention than traces in V'. If V' is empty, we have $V_\pi^\mathsf{C} = V_\pi^{\mathsf{C}'}$. In both cases, PC2 is still satisfied (as π' serves as a contingency for σ in the former case) which concludes this case.

For the second case, assume there exist some $\sigma \in \mathsf{C}$ and $\sigma' \in V_\pi^\mathsf{C}$ such that $\sigma' <_\pi^\mathsf{C} \sigma$ holds. With the same reasoning as before, it follows that $\mathsf{C}' = \mathsf{C} \setminus \mathsf{R}$ with $\mathsf{R} = \{\sigma\}$ is an ω-regular property. Note that $V_\pi^\mathsf{C} = V_\pi^{\mathsf{C}'}$, because the set V' above has to be empty, as there is an intervention trace that is more minimal than σ. Hence PC2 holds for C'. Also, $\pi|_I$ is by definition most minimal, hence $\sigma \neq \pi|_I$ and PC1 holds for C'.

In both cases we have found a smaller ω-regular property $\mathsf{C}' \subset \mathsf{C}$.
"$\Longleftarrow$": By contraposition. Assume that there is some $\mathsf{C}' \subset \mathsf{C}$ that satisfies PC1 and PC2, and let $\sigma \in \mathsf{C} \setminus \mathsf{C}'$. We distinguish between two cases, and show that in any case one of the conjuncts is false.

First, assume $V_\pi^{\mathsf{C}} = V_\pi^{\mathsf{C}'}$. Therefore, as $\sigma \in \mathsf{C}$, we have $\sigma \notin V_\pi^{\mathsf{C}}$ and thus $\sigma \notin V_\pi^{\mathsf{C}'}$. Now, consider all traces σ' such that $\sigma' <_\pi^{\mathsf{C}} \sigma$. There exists at least one such σ' with $\sigma' \notin \mathsf{C}'$, otherwise $\sigma \in V_\pi^{\mathsf{C}'}$ as a minimal intervention trace for C'. Let σ' be such a minimal trace, according to $<_\pi^{\mathsf{C}}$. Now, if σ' was in $\mathsf{C} \setminus \mathsf{C}'$, then $\sigma' \notin V_\pi^{\mathsf{C}}$, but we would have $\sigma' \in V_\pi^{\mathsf{C}'}$ as a minimal intervention, again a contradiction. Therefore, $\sigma' \notin \mathsf{C}$, and since σ' is minimal, we have $\sigma' \in V_\pi^{\mathsf{C}}$. Hence, we have found a σ for which there exists a $\sigma' \in V_\pi^{\mathsf{C}}$ such that $\sigma' <_\pi^{\mathsf{C}} \sigma$, thus the second conjunct is falsified.

For the second case, assume $V_\pi^{\mathsf{C}} \neq V_\pi^{\mathsf{C}'}$. First, consider the case where there is a trace $\sigma' \in V_\pi^{\mathsf{C}'}$ with $\sigma' \notin V_\pi^{\mathsf{C}}$. All traces $\sigma'' <_\pi^{\mathsf{C}} \sigma'$ are in C', and so they are also in C as $\mathsf{C}' \subset \mathsf{C}$. Then, $\sigma' \in \mathsf{C}$, as otherwise, as a minimal intervention for C', it would have also been a minimal intervention for C, and thus in V_π^{C}. Then, since C' is a cause, there exists some contingency π' for σ', $\pi' \in C_\pi^{\sigma'}$ such that $\pi' \not\models \mathsf{E}$, which concludes this case. For the other case, consider $\sigma' \in V_\pi^{\mathsf{C}}$ with $\sigma' \notin V_\pi^{\mathsf{C}'}$. From $\sigma' \in V_\pi^{\mathsf{C}}$ we have $\sigma' \notin \mathsf{C}$ and thus $\sigma' \notin \mathsf{C}'$. Since $\sigma' \notin V_\pi^{\mathsf{C}'}$, there exists a more minimal trace σ'' such that $\sigma'' \in \mathsf{C}$ but $\sigma'' \notin \mathsf{C}'$. Pick σ'' as the most minimal, hence we have $\sigma'' \in V_\pi^{\mathsf{C}'}$. Since C' is a cause, there exists some $\pi'' \in C_\pi^{\sigma''}$ such that $\pi'' \not\models \mathsf{E}$, which concludes this case. In both cases, we have found a trace in C that has a contingency that avoids the effect, which falsifies the first conjunct. □

5.4 Deciding ω-Regular Causality

Putting everything together, we obtain that checking whether some C is a cause for some E on a trace π in system $\mathcal{T}$ can be realized by checking whether the counterfactual automaton $\mathcal{C}_\pi^{\mathcal{T}}$ satisfies a HyperQPTL property, as outlined in the proof of the following theorem.

Theorem 11. *The problem of verifying an ω-regular cause to an ω-regular effect on a lasso-shaped trace is decidable as a HyperQPTL model-checking problem.*

Proof. Assume φ_{C} and φ_{E} are QPTL formulas characterizing the cause and effect properties, respectively, and let π be a lasso-shaped trace. We encode the conditions PC1, PC2 and PC3 directly as a HyperQPTL formula $PC_{\mathsf{E}}^{\mathsf{C}}(\pi)$, utilizing the insight from Lemma 10. The formula is parameterized by π for brevity, however this can be translated to a proper HyperQPTL formula with an additional universal quantifier and a QPTL formula enforcing equality with π. This is possible because π has a lasso shape and can be characterized in QPTL.

In (1), we encode PC1: Both C and E have to be satisfied by the actual trace π. In (2), we enforce that σ is a valid intervention trace with respect to π: All other traces σ' either satisfy C or are not more minimal with respect to π. We then enforce PC2 and PC3. In (4) we state that all traces π' in $\mathcal{C}_\pi^{\mathcal{T}}$ that satisfy

C have to satisfy the effect, and in (3) we enforce that σ is not more minimal than any trace σ'' in C. Together this ensures PC3 due to Lemma 10. The left part of (4) states that there must be a trace π'' under contingency in $\mathcal{C}_\pi^\mathcal{T}$ that violates E with the same input sequence as σ, which corresponds to PC2.

$$
\begin{aligned}
&PC_{\mathsf{E}}^{\mathsf{C}}(\pi) = \forall\sigma\forall\sigma'\forall\sigma''\forall\pi'\exists\pi''.\ \varphi_{\mathsf{C}}(\pi) \land \varphi_{\mathsf{E}}(\pi) \land && (1)\\
&\Big(\neg\varphi_{\mathsf{C}}(\sigma) \land \big(\varphi_{\mathsf{C}}(\sigma') \lor \neg(\psi_{struct}^{\mathsf{C}}(\sigma,\sigma') \land \psi_{min}(\pi,\sigma',\sigma))\big) \rightarrow && (2)\\
&\big((\varphi_{\mathsf{C}}(\sigma'') \rightarrow \neg(\psi_{struct}^{\mathsf{C}}(\sigma'',\sigma) \land \psi_{min}(\pi,\sigma,\sigma''))) \land && (3)\\
&\neg\varphi_{\mathsf{E}}(\pi'') \land \bigwedge_{a\in I} \Box(a_{\pi''} \leftrightarrow a_\sigma))\Big) \land (\varphi_{\mathsf{C}}(\pi') \rightarrow \varphi_{\mathsf{E}}(\pi')) && (4)
\end{aligned}
$$

We then model check the formula $PC_{\mathsf{E}}^{\mathsf{C}}$ against the counterfactual automaton $\mathcal{C}_\pi^\mathcal{T}$. Since model checking HyperQPTL is decidable [26], the theorem follows. □

We conclude by demonstrating the usefulness of an expressive logic such as QPTL for describing causes symbolically, as a similar expression of parity as in the example below would not be possible with previous event-based logics [22].

Example 12. Consider the system $\mathcal{T}$ depicted in Figure 3, the actual trace $\pi = \{\} \cdot (\{a\}\{a,e\})^\omega$, and the effect $\mathsf{E} = \Diamond e$. Disregarding contingencies, the effect can only be avoided by never setting input a at all, i.e., with cause candidate $\mathsf{C}_1 = \Diamond a$ and the resulting set of counterfactual traces $V_\pi^{\mathsf{C}_1} = \{\emptyset^\omega\}$. However, note that the input a at every even position in the trace has no influence on the effect: the system does not discern between the two input sequences $\{\} \cdot (\{a\}\{a\})^\omega$ and $\{\} \cdot (\{a\}\{\})^\omega$. However, that does not mean that the second a is not a potential cause: in the input sequence $\sigma = \{\} \cdot (\{\}\{a\})^\omega$ it is the input that repeatedly moves the trace to state labeled with e. In such situations as on the actual trace π, we say that the second a was *preempted* by the first. Reasoning about contingencies now allows us to find a more accurate cause for E on π. Consider the cause-property $\mathsf{C}_2 = \exists q. \neg q \land \Box(\bigcirc q \leftrightarrow \neg q) \land \Diamond(q \rightarrow a)$, i.e., eventually a holds at an odd position, we have $V_\pi^{\mathsf{C}_2} = \{\sigma\}$. Following the above discussion we have that the trace corresponding to the inputs of σ still satisfies the effect property. However, in the counterfactual automaton we find a trace that agrees with the inputs of σ but avoids the effect: $\pi' = \{\} \cdot (\{\}\{a, e^\mathcal{C}\})^\omega \in \mathit{traces}(\mathcal{C}_\pi^\mathcal{T})$. We have $\pi' \notin \mathsf{E}$. For a short argument why C_2 is also minimal and therefore the cause for E, consider what happens if we require a to appear at multiple (or all) odd positions with $\mathsf{C}_3 = \exists q. \neg q \land \Box(\bigcirc q \leftrightarrow \neg q) \land \Box(q \rightarrow a)$. Now, valid counterfactuals that negate C_3 can be built by simply removing a at some, but not all odd positions, e.g., $\sigma' = \{\}\{a\} \cdot \{\}^\omega \in V_\pi^{\mathsf{C}_3}$. For these sequences, we cannot find a trace in the counterfactual automaton that avoids the effect of the remaining a's at odd positions.

6 Related Work

The increasing number of applications of causality to the formal analysis and explanation of systems has been surveyed comprehensively by Baier et al. [3]. There are several works that define causality formally for computing systems. Gössler and Métayer consider component-based systems and define causality on the component level [12], which differs from our actual causality on the property level. A general framework for counterfactual reasoning in multi-component systems based on counterfactual builders has also been proposed [13], which in particular highlights certain desirable properties of causal analyses. Groce et al. use distance metrics to define the closest trace not producing the effect and define the cause as the difference between the traces [14], which has similarities to our definition of minimal interventions.

Related to our approach, Leitner-Fischer and Leue's causality definitions offer a symbolic description of counterfactual causes in *Event Order Logic* [22]. As effects, they originally considered only violations of safety properties, but their approach has been extended to LTL-definable effects [6]. In both works, the goal of the symbolic causes is to give a high-level description of the orderings of events that lead to a violation in the system, but less to give a precise characterization of the causal input behavior on an observed, actual trace.

Coenen et al. [9] have considered the problem of identifying the actual cause of a counterexample violating a hyperproperty. In their setting, the effect is a hyperproperty while the cause is a concrete set of events appearing in a counterexample. In this work, we consider symbolic causes given as trace properties, and we adapt the counterfactual automata from this aforementioned work [9].

It has been noted before that probabilistic causality can be expressed as a hyperproperty [11,1]. The considered version of probabilistic causation is founded on the probability raising principle. However, this type of probabilistic causation can also be expressed in branching-time temporal logics, as shown by Kleinberg and Mishra [20]. For probabilistic systems, there has recently been proposed a notion of causality that combines probability raising with the counterfactuality principle [28]. To the best of our knowledge, the observation that counterfactual causality is not a trace property [10] has not been formalized and proven before.

7 Conclusion

Inspired by Halpern and Pearl's definition of actual causality, we define causality for reactive systems that gives symbolic descriptions of causal temporal behavior as trace properties. We define interventions and contingencies to enable counterfactual reasoning in this infinite setting. The key idea of our work is to link satisfaction of a property with a distance measure over traces, to define the closest counterfactual traces that do not satisfy the cause. We show that checking causality for trace properties cannot itself be expressed as a trace property but as a hyperproperty. Our definitions can be instantiated with explicit logics to express cause and effect properties. We present a decidable instantiation with

QPTL along with the corresponding automata-based constructions to verify actual causes based on HyperQPTL model checking, covering the whole practically relevant class of ω-regular properties. Future work includes examining ways of leveraging the existing research on hyperproperties when analyzing causal relationships, and applying our conceptual framework to other domains.

References

1. Ábrahám, E., Bonakdarpour, B.: Hyperpctl: A temporal logic for probabilistic hyperproperties. In: QEST 2018
2. Baier, C., Coenen, N., Finkbeiner, B., Funke, F., Jantsch, S., Siber, J.: Causality-based game solving. In: CAV 2021
3. Baier, C., Dubslaff, C., Funke, F., Jantsch, S., Majumdar, R., Piribauer, J., Ziemek, R.: From Verification to Causality-Based Explications. In: ICALP 2021
4. Beer, I., Ben-David, S., Chockler, H., Orni, A., Trefler, R.: Explaining counterexamples using causality. Formal Methods in System Design (2012)
5. Buechi, J.R.: On a decision method in restricted second-order arithmetic. In: International Congress on Logic, Methodology, and Philosophy of Science (1962)
6. Caltais, G., Guetlein, S.L., Leue, S.: Causality for general ltl-definable properties. In: CREST@ETAPS 2018
7. Chockler, H., Halpern, J.Y., Kupferman, O.: What causes a system to satisfy a specification? ACM Trans. Comput. Log. (2008)
8. Clarkson, M.R., Schneider, F.B.: Hyperproperties. J. Comput. Secur. (2010)
9. Coenen, N., Dachselt, R., Finkbeiner, B., Frenkel, H., Hahn, C., Horak, T., Metzger, N., Siber, J.: Explaining Hyperproperty Violations. In: CAV 2022
10. Datta, A., Garg, D., Kaynar, D.K., Sharma, D., Sinha, A.: Program actions as actual causes: A building block for accountability. In: CSF 2015
11. Dimitrova, R., Finkbeiner, B., Torfah, H.: Probabilistic hyperproperties of markov decision processes. In: ATVA 2020
12. Gössler, G., Métayer, D.L.: A general trace-based framework of logical causality. In: FACS 2013
13. Gössler, G., Stefani, J.: Causality analysis and fault ascription in component-based systems. Theor. Comput. Sci. (2020)
14. Groce, A., Chaki, S., Kroening, D., Strichman, O.: Error explanation with distance metrics. Int. J. Softw. Tools Technol. Transf. (2006)
15. Halpern, J.Y.: A modification of the halpern-pearl definition of causality. In: IJCAI 2015
16. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach. part i: Causes. The British Journal for the Philosophy of Science (2005)
17. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach. part ii: Explanations. The British Journal for the Philosophy of Science (2005)
18. Holzmann, G.J.: The model checker SPIN. IEEE Trans. Software Eng. (1997)
19. Hume, D.: An Enquiry Concerning Human Understanding. London (1748)
20. Kleinberg, S., Mishra, B.: The temporal logic of causal structures. In: UAI 2009
21. Kupriyanov, A., Finkbeiner, B.: Causal termination of multi-threaded programs. In: CAV 2014
22. Leitner-Fischer, F., Leue, S.: Causality checking for complex system models. In: VMCAI 2013
23. Lewis, D.K.: Counterfactuals. Cambridge, MA, USA: Blackwell (1973)

24. Moore, E.F.: Gedanken-experiments on sequential machines. Aut. stud. (1956)
25. Pnueli, A.: The temporal logic of programs. In: FOCS 1977
26. Rabe, M.N.: A temporal logic approach to information-flow control. Ph.D. thesis, Saarland University (2016)
27. Sistla, A.P.: Theoretical Issues in the Design and Verification of Distributed Systems. Ph.D. thesis (1983)
28. Ziemek, R., Piribauer, J., Funke, F., Jantsch, S., Baier, C.: Probabilistic causes in markov chains. Innov. Syst. Softw. Eng. (2022)