The Case for Time in Causal DAGs

Alexander G. Reisach[1]
CNRS, MAP5
Université Paris Cité
F-75006 Paris, France

Alberto Suárez
Departamento de Ingeniería Informática,
Escuela Politécnica Superior,
Universidad Autónoma de Madrid, Spain

Sebastian Weichwald
Department of Mathematical Sciences and
Pioneer Centre for Artificial Intelligence,
University of Copenhagen, Denmark

Antoine Chambaz
CNRS, MAP5
Université Paris Cité
F-75006 Paris, France

Abstract

We make the case for incorporating time explicitly into the definition of variables in causal directed acyclic graphs (DAGs). Causality requires that causes precede effects in time, meaning that the causal relationships between variables in one time order may not be the same in another. Therefore, any causal model requires temporal qualification; this applies even if the model does not describe a time series of repeated measurements. We formalize a notion of time for causal variables and argue that it resolves existing ambiguity in causal DAGs and is essential to assessing the validity of the acyclicity assumption. If variables are separated in time, their causal relationship is necessarily acyclic. Otherwise, acyclicity depends on the absence of any causal cycles permitted by the time order. We introduce a formal distinction between these two conditions and lay out their respective implications. We outline connections of our contribution with different strands of the broader causality literature and discuss the ramifications of considering time for the interpretation and applicability of DAGs as causal models.

Prologue

DEMOCRITUS: I would rather discover one cause than gain the kingdom of Persia.
SOCRATES: Say you discover that the pursuit of knowledge causes subsequent happiness.
DEMOCRITUS: Very well.
SOCRATES: Is it possible that later happiness might be the cause of past pursuit of knowledge?
DEMOCRITUS: Impossible!
SOCRATES: But perhaps intermittent happiness may cause continued pursuit of knowledge?
DEMOCRITUS: Likely so.
SOCRATES: Between the pursuit of knowledge and happiness, which one is the cause again?
DEMOCRITUS: ... there is a kingdom-merchant I may need to speak to.

[1] Correspondence to alexander.reisach@math.cnrs.fr

1 Introduction

Causal directed acyclic graphs (DAGs), as popularized by Pearl (2000), have been conceived as nontemporal causal models. In the second edition of the same work, Pearl (2009), Chapter 7.5.1 states on the matter of using time to determine causation:

> "The reliance on temporal information has its price, as it excludes a priori the analysis of cases in which the temporal order is not well-defined."

Time is real, however, and there is no escaping it and whatever price it really may exact, whether temporal information is relied upon or not. We therefore argue that it is best to use time explicitly, and set out to explore the limitations and opportunities it reveals.

Time and causation. An essential notion that we rely on throughout this work is the idea that *causes precede effects*. This notion is fundamental to the definition of causality. In proposing the method of path coefficients, the precursor to causal graphical models, Wright (1934) states that "only such paths are tried which are appropriate in direction in time", which ensures that the paths form a DAG. The Bradford Hill criteria for causality (Hill 1965), widely used in epidemiology, refer to the same principle as 'temporality' and posit it as one of nine criteria for establishing evidence of causation. Similarly, Granger (1969), Suppes (1970), and Shoham (1990) rely on the time order between cause and effect for their theories of causality. Throughout their foundational work on causal discovery, Spirtes et al. (1993) (see also Spirtes et al. 2001, for the second edition) emphasize the existence of a time order which ought not to be violated by causal models and ensures the uniqueness of the DAG. Their focus on causal reasoning without time structure notwithstanding, the canonical work on causal graphical models Pearl (2009), e.g. Chapter 7.5.1, as well as Peters et al. (2017), e.g. Chapter 10, concur that causality must adhere to the time order.

The absence of time in causal DAGs. Despite its central role in causality, the notion of time is absent from the formal treatment of causal DAGs in the existing literature. In contrast to other approaches to causal analysis, in particular those based on the potential outcomes framework (see e.g. Imbens and Rubin 2015), causal DAGs have been developed explicitly for 'nontemporal statistical data' (Pearl 2009, Chapter 2). This expression covers two distinct properties. First, every DAG is stationary in the sense that any distribution that is compatible with it does not depend on time. Second and perhaps more controversially, there are no explicit temporal conditions among the components of an individual sample from such a distribution beyond those entailed by the causal order imposed by the acyclicity of the DAG. One can think of the first property as 'between-sample', and of the second one as 'within-sample' (see forthcoming Figure 1 for an illustration). Instead of relying on a time order for the existence of a causal order, most of the literature follows Pearl (2009) in assuming that the causal variables are Markov with respect to a DAG. This offers a powerful unified theory regardless of whether the causal order of a DAG is due to a time order or the absence of cyclic causal relationships (Pearl 2009, Section 7.5.1). However, the general nature of this approach obscures the need for temporal precedence in causation and the rationale for why the acyclicity assumption may or may not be valid.

As a result, causal DAGs have been used widely, including in contexts such as neuroimaging (Ramsey et al. 2017) or gene regulatory networks (Lee et al. 2019), where the existence of a causal

DAG is questionable. In many cases, the search for nontemporal causal relationships has lead to variables that do not easily allow for well-defined interventions, such as the Altitude $\to$ Temperature example (see e.g. Peters et al. 2017, Section 2.1). The absence of time has also prompted the use of time proxy variables like 'age' or 'hour of day' (Mooij et al. 2016), and the modelling of time itself as a cause (Huang et al. 2015).

The case for time in causal DAGs. We make the case for a formal and explicit treatment of the time of variables in causal DAGs for two main reasons. First, because time is essential for the interpretation of causal relationships since causes must precede effects. For example, the statement "X does not cause Y", for two causal variables X and Y, is meaningful if X precedes Y, but is a technicality if the time order were opposite. Yet, without a notion of time, the two cases are indistinguishable in a causal DAG (note that DAGs cannot express the absence of causal knowledge as outlined by Dawid 2010, Sections 9 and 10). Second, because time is crucial to ensuring that the acyclicity assumption underlying the model class is valid. For example, there is certain to be a (possibly empty) DAG between two causal variables X and Y if one precedes the other, but there may not be a DAG without such a time order, since they could each influence the other in a cyclic fashion. We emphasize that these reasons apply generally, not only to time series of repeated measurements. An explicit notion of time offers to resolve ambiguities regarding the causal relationships intended to be represented by DAGs, and is crucial for assessing the validity of the acyclicity assumption.

Contribution. We begin by noting that causal relationships are relative with respect to time and propose a formalism that incorporates time into the definition of causal variables. On this basis, we characterize two distinct and individually sufficient conditions for acyclicity: acyclicity by a time order that prevents cycles, and acyclicity by the absence of causal cycles where the time order does not prevent them. We describe their implications, in particular how they differ for evaluating the validity of the acyclicity assumption, and show that they can help uncover faithfulness violations. We outline connections of our contribution to critical perspectives on causal DAGs, cyclic causal graphs, discrete and continuous time causal models, and the potential outcomes framework. Finally, we discuss the ramifications of considering time in DAGs for their interpretation and applicability as causal models.

2 Time in Causal DAGs

We motivate and formalize a notion of time and time ordering for the variables in causal DAGs.

Causal DAGs. DAGs are the central tool for representing causal relationships using graphical models (Pearl 2009). A DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of vertices $\mathcal{V}$, and a set of directed edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, such that if $(X, Y) \in \mathcal{E}$, then there is a directed edge from X to Y. We call 'directed path' an ordered collection of directed edges such that the end point of the k-th edge is the start point of the $(k+1)$-th edge. By definition, DAGs are acyclic, that is, there are no directed paths that begin and end in the same node. A causal DAG represents an external causal system (Dawid 2010, Section 5). Each vertex corresponds to a random variable, and their joint distribution is assumed to be Markov with respect to the DAG (e.g. Pearl 2009, Definition 1.2.2; Peters et al. 2017, Definition 6.21). A directed edge $(X, Y) \in \mathcal{E}$, also denoted as $X \to Y$, indicates that X is a direct cause of Y within the variables of the graph. In the literature on causal graphical models, 'causal structure' and 'causal model' are defined as a DAG, and a DAG with a set of additional parameters, respectively (Pearl 2009, Definitions 2.2.1 and 2.2.2). The superclass of 'structural causal models' (Pearl 2009, Definition 7.1.1) does not in principle require acyclic structures, but is nonetheless most often used in conjunction with DAGs (see e.g. Pearl 2009, Chapter 1.2; Peters et al. 2017, Chapter 6.2), which remain the central object of study and interest in the field of causal graphical models due to their favorable properties.

2.1 Causal Relationships are Relative With Respect to Time

Since causes must precede effects, causal relationships are inevitably dependent on the temporal relationship between variables. There can be no causal relationship without temporal qualification.

A time-relative causal relationship. For an example of the relative nature of causal relationships with respect to time, consider the DAG Aspirin → Headache (a popular example in the literature, e.g. Shpitser and Pearl 2008; Imbens and Rubin 2015; Pearl and Mackenzie 2018).[2] Clearly, Aspirin cannot relieve a past Headache, so the effect is time-specific. Moreover, may Headache not cause subsequent intake of Aspirin? It is not clear if the absent edge from Headache to Aspirin is specific to the time order, or if it is of a general nature. The presence of the causal edge implies a time order of Aspirin before Headache, or, if each variable corresponds to multiple time points, that at least some temporal component of Aspirin precedes some temporal component of Headache, but it does not specify the temporal relationship exactly. A causal DAG consisting of two disconnected variables would give no indication whatsoever about the time order, and the interpretation of absent edges would be wholly ambiguous. In the example, the ambiguity can obviously be resolved by instantiating the concepts of Aspirin and Headache as time-specific variables such as 'Aspirin in the morning' and 'Headache in the afternoon', with an edge pointing from the former to the latter. However, there is most likely also an edge pointing from 'Headache in the morning' to 'Aspirin in the afternoon', meaning that if both time points were lumped back together into the concepts of Aspirin and Headache, there would be no DAG at all since the two edges would form a cycle. The example highlights that a DAG cannot, in fact, relate the general concepts of Aspirin and Headache, but time-specific instantiations thereof. Such instantiations are standard in the potential outcomes literature (see e.g. Imbens and Rubin 2015, Chapter 1). Acknowledging the relativity of causal relationships with respect to time thus reveals what kind of variables can be related by causal DAGs. As the example shows, the variables in causal DAGs, as well as the causal relationship between them, may be of a rather specific nature.

Time at the level of individual samples. The majority of the literature on causal graphical models follows Pearl in employing the conception of Laplace (1814) that natural phenomena can be understood as the outputs of deterministic transformations applied to random inputs (see e.g. Pearl 2009, Section 1.4). The transformations constitute the causal mechanisms. Assuming some random inputs are mutually independent, it follows that the corresponding outputs are mutually independent as well. Note the contrast to time-series causality (see e.g. Assaad 2021), where the output of a causal mechanism at one time step affects its input at the next time step, leading to dependence. Nonetheless,

[2] Aspirin is a drug frequently used for headache relief.

even in the case of mutually independent inputs, the steps of a causal mechanism take place in time within every application. Thus, elements of an output must correspond to one or multiple points in time. Peters et al. (2017), state in Remark 6.7 that "in the [structural causal model], the variables no longer describe measurements at specific points in time". Our work concerns the relative time delta between elements of the output, which must exist if causes are to precede effects in time.

An output can be seen as an individual sample from the distribution given by applying a causal mechanism to random inputs. The time delta between elements of one output is within-sample. The time delta delta between elements of different outputs, on the other hand, is between-sample. For simplicity, and to be independent of the start time, we assume the earliest time point for each individual sample to be 0, which can be achieved by subtracting the earliest time point per individual sample. Figure 1 provides a schematic illustration of within-sample time and between-sample time.

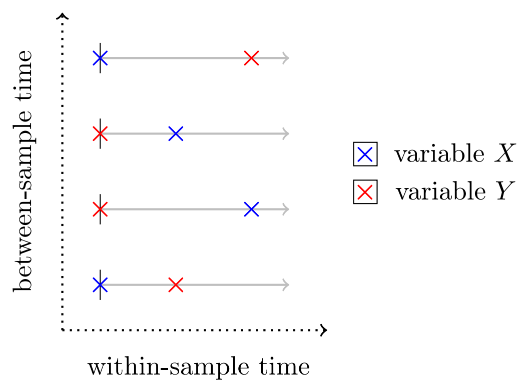


Figure 1: Schematic of within-sample time and between-sample time for four samples of two variables X (blue) and Y (red). Each individual sample spawns its own within-sample timeline (solid arrows), along which causal effects take place and its variables are measured from left to right. Time may also pass between-sample in the vertical direction from bottom to top.

2.2 A Formal Notion of Time for Causal Variables

Since causal relationships are relative with respect to time, no single DAG can describe causal relationships in a time-invariant fashion (except in the trivial case of an empty DAG denoting no causal effects regardless of time). In order to speak of a unique and non-empty DAG between variables, it is necessary to endow the variables with a suitable notion of time. To this end, we introduce a formal treatment of within-sample time for causal variables. For convenience, we will use the following notation: for any $k \in \mathbb{N}^*$, for any set S of cardinality greater than or equal to k, $\mathcal{P}_k(S)$ denotes the collection of subsets of S which are of cardinality k.

Let $\mathcal{T}_x \subset \mathbb{R}_+$ be a countable set of time points, and let $k_x \in \mathbb{N}^*$ be a deterministic integer smaller than or equal to $|\mathcal{T}_x|$. We define a real-valued stochastic process $\mathbb{X} = (\mathbb{X}_t)_{t \in \mathcal{T}_x}$ over these time points, and denote by $\mathbb{T}_x \subseteq \mathcal{T}_x$ a random subset consisting of k_x time points. The support $\mathcal{S}_x$ of the distribution of $\mathbb{T}_x$ is a subset of $\mathcal{P}_{k_x}(\mathcal{T}_x)$. The variables $\mathbb{X}_t$ are measured for every $t \in \mathbb{T}_x$.

Let $f_x \colon \mathbb{R}^{k_x} \to \mathbb{R}$ be a real-valued deterministic aggregation function, so that $X = f_x((\mathbb{X}_t)_{t \in \mathbb{T}_x})$ is an aggregate variable of the measurements $(\mathbb{X}_t)_{t \in \mathbb{T}_x}$, which form its temporal components. We impose further that if $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{k_x}$ are such that there exists $t \in \{1, \ldots, k_x\}$ for which $\mathbf{x}_s = \mathbf{x}'_s$ for every $s \neq t$ but $\mathbf{x}_t \neq \mathbf{x}'_t$, then $f_x(\mathbf{x}) \neq f_x(\mathbf{x}')$. This is to ensure that an intervention on one temporal component also results in a change in the corresponding aggregate variable. For instance, f_x cannot act as a component selector and it cannot multiply variables if one of them is equal to zero.

To emphasize how the aggregate variable X inherits its temporal structure from the underlying stochastic process, we introduce the notation

$$X_{\mathbb{T}_x} \coloneqq f_x\left((\mathbb{X}_t)_{t\in\mathbb{T}_x}\right), \tag{1}$$

which captures that a causal variable can be seen as the aggregate of one or multiple temporal components. Owing to the essential role of time in causality, we choose to keep the time index explicit even after the aggregation into a single real-valued random variable. For a minimal example, we can choose $\mathcal{T}_x = \{t\}$, so that the stochastic process $\mathbb{X}$ consists of a single variable, and $k_x = 1$ with f_x the identity, so that $X_{\mathbb{T}_x} = \mathbb{X}_t$ (forthcoming examples focus on processes of multiple variables to illustrate the properties of our formalism). In summary, we formalize a causal variable $X_{\mathbb{T}_x}$ in terms of the objects contained in Table 1.

$X_{\mathbb{T}_x}$	Causal variable with an explicit notion of time.
$\mathbb{X}$ or $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$	Stochastic process of measurements.
$\mathcal{T}_x$	Set of possible measurement time points.
$\mathbb{T}_x$	Random subset of k_x time points from $\mathcal{T}_x$ used for $X_{\mathbb{T}_x}$.
k_x	Number of measurements per individual sample of $X_{\mathbb{T}_x}$.
f_x	Aggregation function of $(\mathbb{X}_t)_{t\in\mathbb{T}_x}$.

Table 1: The objects used for formalizing a causal variable $X_{\mathbb{T}_x}$ with an explicit notion of time.

Moving forward, all causal variables will be of the form given in Equation (1), and we assume the existence of an underlying 'atomic' causal DAG between all time point-specific components of the stochastic processes involved in the phenomenon of interest. From the atomic causal DAG it is possible to derive a directed graph between these aggregate variables, which may or may not be acyclic. For a detailed treatment of inference rules over macro-variables, see Anand et al. (2023).

By default, we take the set of time points $\mathbb{T}_x$ of a variable $X_{\mathbb{T}_x}$ to be the points in time at which $\mathbb{X}$ is measured, but other specifications are also possible (e.g. proxy measurements of past events or measurements representative of multiple time points). The scope of interventions on a causal variable $X_{\mathbb{T}_x}$ is given by the interventions on any subset of $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$. Denoting the support of the joint distribution of two sets of time points $(\mathbb{T}_x, \mathbb{T}_y)$ as

$$\mathcal{S}_{xy} \subseteq \mathcal{S}_x \times \mathcal{S}_y \subseteq \mathcal{P}_{k_x}(\mathcal{T}_x) \times \mathcal{P}_{k_y}(\mathcal{T}_y), \tag{2}$$

we say that $X_{\mathbb{T}_x}$ causes another causal variable $Y_{\mathbb{T}_y}$ if there exist $(t_x, t_y) \in \mathcal{S}_{xy}$ such that $\mathbb{X}_{t_x}$ causes $\mathbb{Y}_{t_y}$ in the atomic DAG (we do not consider self-loops on the level of aggregate variables, e.g. if $\mathbb{X}_{t_{x_1}}$ causes $\mathbb{X}_{t_{x_2}}$ for some $t_{x_1}, t_{x_2} \in \mathcal{S}_x$).

Example 1 (separation in time). Consider a generic medical treatment-health setting following the formalization introduced above. Let $X_{\mathbb{T}_x}$ be an aggregate of $k_x = 2$ treatment measurements in $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$, where $\mathcal{T}_x = \{0, 0.4\}$. Let $Y_{\mathbb{T}_y}$ be an aggregate variable of $k_y = 2$ health measurements in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, where $\mathcal{T}_y = \{0.6, 1\}$. It follows that $\mathbb{T}_x = \mathcal{T}_x$ and $\mathbb{T}_y = \mathcal{T}_y$. Thus, the only possible combination of time points is $\mathcal{S}_{xy} = \mathcal{P}_2(\mathcal{T}_x) \times \mathcal{P}_2(\mathcal{T}_y) = \{(\{0, 0.4\}, \{0.6, 1\})\}$. Let the aggregation functions f_x and f_y take the sum of their inputs, that is, $X_{\mathbb{T}_x} \coloneqq \sum_{t \in \mathbb{T}_x} \mathbb{X}_t$ and $Y_{\mathbb{T}_y} \coloneqq \sum_{t \in \mathbb{T}_y} \mathbb{Y}_t$. Such a setting may correspond to patients always receiving a treatment at both of two possible time points, followed always by a health screening at both of two possible time points.

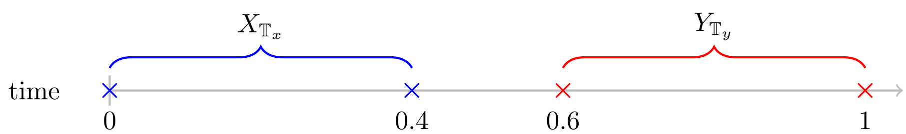
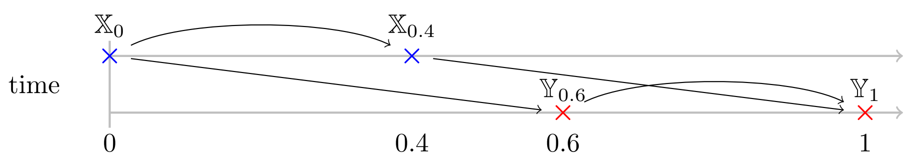


Figure 2: Two aggregate causal variables that are separated in time.

Figure 2 provides a visual illustration of the example, including the timeline, the time points of the stochastic process $(\mathbb{X}_t)_{t \in \mathcal{T}_x}$ (blue $\times$) and $(\mathbb{Y}_t)_{t \in \mathcal{T}_y}$ (red $\times$), and their aggregation into the corresponding aggregate variables $X_{\mathbb{T}_x}$ (blue bracket) and $Y_{\mathbb{T}_y}$ (red bracket). The aggregate variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ can be thought of as time-specific instantiations of the abstract concepts 'treatment' and 'health'. Any causal relationships between them are by default specific to this instantiation.



(a) Atomic causal DAG.

$$X_{\mathbb{T}_x} \longrightarrow Y_{\mathbb{T}_y}$$

(b) Causal DAG between the aggregate variables.

Figure 3: An atomic causal DAG and corresponding aggregate causal DAG compatible with Example 1.

Figure 3 shows an atomic causal DAG (Figure 3a) that is compatible with the example, and its corresponding causal DAG between the aggregate variables (Figure 3b). The temporal components of each aggregate variable are drawn on separate (gray) time arrows belonging to the same within-sample timeline, as is indicated by the vertical connection. The solid black arrows show causal relationships.

Using our formalization, we can define the notion that causes must precede effects rigorously. Let a wavy arrow $\rightsquigarrow$ denote a causal path, whether direct or indirect, between two variables within a causal directed graph, acyclic or not. If the causal path consists of a single edge, then $\rightsquigarrow$ can be replaced by a straight arrow $\rightarrow$.

Definition 1 (Causes precede effects)**.** *We say that causes precede effects if and only if, in the atomic causal DAG between the temporal components of any pair of stochastic processes* $(\mathbb{X}_t)_{t \in \mathcal{T}_x}$ *and* $(\mathbb{Y}_t)_{t \in \mathcal{T}_y}$*,*

$$\forall t_x \in \mathcal{T}_x, \forall t_y \in \mathcal{T}_y, \left(\mathbb{X}_{t_x} \rightsquigarrow \mathbb{Y}_{t_y} \implies t_x < t_y\right).$$

This allows verifying the basic requirement that causes precede effects for variables of the form given in Equation (1). For example, a path $X_{\mathbb{T}_x} \rightsquigarrow Y_{\mathbb{T}_y}$ in Example 1 would respect Definition 1, but the inverse $Y_{\mathbb{T}_y} \rightsquigarrow X_{\mathbb{T}_x}$ would not. Throughout this work, we presume that all causal relationships respect that causes precede effects by Definition 1. Building on this, our formalization of the time associated with causal variables enables us to determine the possibility of causal cycles and to state precise requirements for acyclicity in terms of the atomic causal DAG.

3 A Tale of Two Acyclicities

We partition the assumption of acyclicity into two separate and individually sufficient conditions with distinct characteristics: acyclicity by a time order that prevents cycles, and acyclicity by the absence of causal cycles where the time order does not prevent them.

Acyclicity of the causal relationship between two variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ can be defined in terms of the causal relationships between the temporal components of the variables, that is, the atomic DAG.

Definition 2 (Acyclicity)**.** *Acyclicity between two causal variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ holds if and only if in the atomic causal DAG*

$$\exists (t_x, t_y) \in \mathcal{S}_x \times \mathcal{S}_y \text{ such that } \mathbb{X}_{t_x} \rightsquigarrow \mathbb{Y}_{t_y} \implies \forall (t'_x, t'_y) \in \mathcal{S}_x \times \mathcal{S}_y, \neg \left(\mathbb{Y}_{t'_y} \rightsquigarrow \mathbb{X}_{t'_x} \right).$$

We further call a graph $\mathcal{G}$ acyclic if acyclicity holds for all pairs of variables in $\mathcal{G}$.

3.1 Acyclicity I: By Time Order

A sufficient condition for an acyclic causal relationship between two variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ is satisfied if all time points of one variable always precede those of the other variable. We denote the temporal precedence of $X_{\mathbb{T}_x}$ with respect to $Y_{\mathbb{T}_y}$ by the shorthand

$$\mathbb{T}_x \prec \mathbb{T}_y \coloneqq \forall (s_x, s_y) \in \mathcal{S}_{xy}, \max(s_x) < \min(s_y) \tag{3}$$

(the order relation $\prec$ is transitive), and define ‘time-acyclicity’ as follows.

Definition 3 (Time-acyclicity)**.** *Time-acyclicity between two causal variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ holds if and only if*

$$(\mathbb{T}_x \prec \mathbb{T}_y) \vee (\mathbb{T}_y \prec \mathbb{T}_x).$$

We further call a graph $\mathcal{G}$ time-acyclic if time-acyclicity holds for all pairs of variables in $\mathcal{G}$.

The restriction on $\mathcal{S}_{xy}$ posed by Definition 3 trivially implies acyclicity by Definition 2; this is particularly easy to see in the case of two variables of the form given in Equation (1) with a single measurement time point each, and hence the same causal DAG as the atomic causal DAG.

Example 1 (separation in time) revisited. Example 1 exhibits time-acyclicity by Definition 3, as can be seen by the horizontal separation of blue and red brackets along the time axis. The aggregate treatment variable $X_{\mathbb{T}_x}$ could be a cause of the aggregate health variable $Y_{\mathbb{T}_y}$, but not vice versa due to the time order $X_{\mathbb{T}_x} \prec Y_{\mathbb{T}_y}$. A path $X_{\mathbb{T}_x} \rightsquigarrow Y_{\mathbb{T}_y}$ indicates that at least one of the two measured treatments causes at least one of the two measured health outcomes, as is the case in the atomic causal DAG shown in Figure 3.

Example 2 (non-obvious time-acyclicity). Consider a variation of the generic treatment-health setting of Example 1. Let $Z_{\mathbb{T}_z}$ be an aggregate of $k_z = 1$ diagnostic measurement in $(\mathbb{Z}_t)_{t \in \mathcal{T}_z}$, where $\mathcal{T}_z = \{0\}$. Let $X_{\mathbb{T}_x}$ be an aggregate of $k_x = 1$ treatment measurement in $(\mathbb{X}_t)_{t \in \mathcal{T}_x}$, where $\mathcal{T}_x = \{0.25, 0.7\}$. Let $Y_{\mathbb{T}_y}$ be an aggregate of $k_y = 1$ health measurement in $(\mathbb{Y}_t)_{t \in \mathcal{T}_y}$, where $\mathcal{T}_y = \{0.6, 1\}$. Let the aggregation functions f_z, f_x, and f_y each be the identity. Let further $\mathcal{S}_{xy} = \{(\{0.25\}, \{0.6\}), (\{0.25\}, \{1\}), (\{0.7\}, \{1\})\}$. Note that $(\{0.7\}, \{0.6\}) \notin \mathcal{S}_{xy}$, meaning that $X_{\mathbb{T}_x} \prec Y_{\mathbb{T}_y}$ despite $\max(\mathcal{T}_x) > \min(\mathcal{T}_y)$. Such a setting may correspond to patients undergoing a diagnostic screening, a subsequent treatment at one of two time points, and a final health screening at one of two other time points, but always after the treatment.

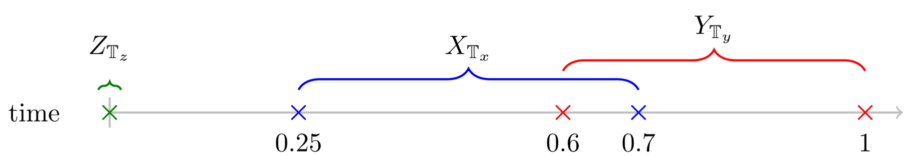


Figure 4: An illustration of Example 2, where the time order prevents causal cycles. The acyclicity assumption can be validated on the basis of time-acyclicity (Definition 3), but doing so requires a detailed understanding of the joint distribution of measurement times.

The example is illustrated in Figure 4. Although it may seem as if there were no time order between $X_{\mathbb{T}_x}$ (blue) and $Y_{\mathbb{T}_y}$ (red) due to the overlap of the brackets connecting their respective time points, the joint distribution of $\mathbb{T}_x$ and $\mathbb{T}_y$ actually guarantees $X_{\mathbb{T}_x} \prec Y_{\mathbb{T}_y}$, meaning that there is time-acyclicity. The variable $Z_{\mathbb{T}_z}$ gives a reference time $\mathbb{T}_z$ with respect to which there can be $t_x > t_y$ with $t_x \in \mathbb{T}_x$ and $t_y \in \mathbb{T}_y$, provided t_x and t_y correspond to different samples, but not within the same sample. As the example shows, ensuring time-acyclicity may require a detailed understanding of the measurement procedure.

3.2 Acyclicity II: By Absence of Causal Cycles

Discussions of the acyclicity assumption in the causal DAG literature typically focus on the presence or absence of causal cycles such as feedback loops (see e.g. Brouillard et al. 2024). However, given that causes precede effects, causal cycles can only occur if there is no time-acyclicity (Definition 3). We term acyclicity due to the absence of causal cycles where they are possible 'effect-acyclicity' and define it as follows.

Definition 4 (Effect-acyclicity)**.** *Effect-acyclicity holds between two causal variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ if and only if there is no time-acyclicity by Definition 3, that is,*

$$\neg \left((\mathbb{T}_x \prec \mathbb{T}_y) \vee (\mathbb{T}_y \prec \mathbb{T}_x) \right),$$

yet there is acyclicity by Definition 2. We further call a graph $\mathcal{G}$ effect-acyclic if effect-acyclicity holds for all pairs of variables in $\mathcal{G}$.

Example 3 (effect-acyclicity). Consider another variation of our running medical treatment-health example. Let $X_{\mathbb{T}_x}$ be an aggregate of $k_x = 2$ treatment measurements in $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$ where $\mathcal{T}_x = \{0, 0.6\}$. Let $Y_{\mathbb{T}_y}$ be an aggregate of $k_y = 2$ health measurements in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, where $\mathcal{T}_y = \{0.4, 1\}$. It follows that $\mathbb{T}_x = \mathcal{T}_x$, $\mathbb{T}_y = \mathcal{T}_y$, and $\mathcal{S}_{xy} = \mathcal{P}_2(\mathcal{T}_x) \times \mathcal{P}_2(\mathcal{T}_y) = \{(\{0, 0.6\}, \{0.4, 1\})\}$. Let the aggregation functions f_x and f_y each take the sum of their inputs. In this example, there is no time-acyclicity, meaning that a causal cycle between the aggregate causal variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ is possible. This would be the case for example in an adaptive treatment regime where a patient receives treatment at time $t = 0$, which affects health at time 0.4, which affects treatment at time 0.6, which in turn affects health at time 1.

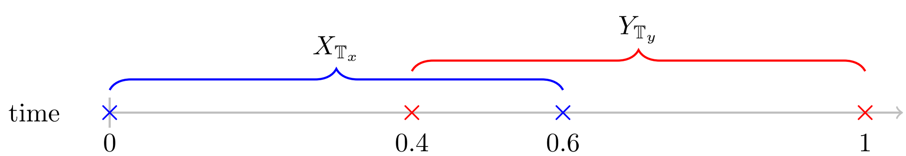


Figure 5: An illustration of Example 3, where the time order permits causal cycles. Therefore, the acyclicity assumption can only hold by effect-acyclicity (Definition 4).

An illustration of the example is shown in Figure 5. The overlap of the red and blue brackets indicates a lack of time order by Equation (3) between $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$, as can be read off $\mathcal{S}_{xy}$. The acyclicity assumption can only hold by effect-acyclicity (Definition 4), otherwise the variables cannot be validly represented in a causal DAG. Provided there is no causal cycle, the time order allows for any causal relationship between $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$. A causal DAG including the path $X_{\mathbb{T}_x} \rightsquigarrow Y_{\mathbb{T}_y}$ would require that $\mathbb{Y}_{0.4}$ did not cause $\mathbb{X}_{0.6}$. Analogously, a causal DAG including the path $Y_{\mathbb{T}_y} \rightsquigarrow X_{\mathbb{T}_x}$ would require that $\mathbb{X}_0$ did not cause either $\mathbb{Y}_{0.4}$ or $\mathbb{Y}_1$, and that $\mathbb{X}_{0.6}$ did not cause $\mathbb{Y}_1$. As the example shows, effect-acyclicity relies on specific causal relationships in the atomic causal DAG between the temporal components of the different variables.

Effect-acyclicity regardless of time. A special case of effect-acyclicity between two causal variables is given if there is no way for one to affect the other, regardless of the temporal relationship between them. We refer to this condition as 'total effect-acyclicity' and define it as follows.

Definition 5 (Total effect-acyclicity)**.** *Total effect-acyclicity between two real-valued stochastic processes $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$ and $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$ holds if and only if in the atomic causal DAG*

$$\exists (t_x, t_y) \in \mathcal{T}_x \times \mathcal{T}_y \textit{ such that } \mathbb{X}_{t_x} \rightsquigarrow \mathbb{Y}_{t_y} \implies \forall (t'_x, t'_y) \in \mathcal{T}_x \times \mathcal{T}_y, \neg\left(\mathbb{Y}_{t'_y} \rightsquigarrow \mathbb{X}_{t'_x}\right).$$

In this case, we say that total effect-acyclicity holds for any corresponding $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$. We further call a graph $\mathcal{G}$ totally effect-acyclic if effect-acyclicity holds for all pairs of variables in $\mathcal{G}$.

Total effect-acyclicity by Definition 5 can be seen as an extension of acyclicity by Definition 2 to all time points of the stochastic processes.

Example 4 (total effect-acyclicity). Consider a hypothetical causal model between solar flares and power outages. Assume that solar flares can cause power outages but power outages cannot physically cause solar flares. Let $X_{\mathbb{T}_{x_1}}$ be an aggregate of $k_{x_1} = 2$ solar flare measurements in $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$, where $\mathcal{T}_x = \{0, 0.2, 0.8, 1\}$, such that $\mathcal{S}_{x_1} = \{(\{0, 0.2\})\}$. Let further $X_{\mathbb{T}_{x_2}}$ be an another aggregate of $k_{x_2} = 2$ solar flare measurements in $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$, such that $\mathcal{S}_{x_2} = \{(\{0.8, 1\})\}$. Let $Y_{\mathbb{T}_y}$ be an aggregate variable of $k_y = 2$ power outage measurements in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, where $\mathcal{T}_y = \{0.4, 0.6\}$, meaning that $\mathbb{T}_y = \mathcal{T}_y$. It follows that $S_{x_1 y} = \{(\{0, 0.2\}, \{0.4, 0.6\})\}$ and that $\mathcal{S}_{yx_2} = \{(\{0.4, 0.6\}, \{0.8, 1\})\}$. Let the aggregation functions f_{x_1}, f_{x_2}, and f_y each take the sum of their inputs.

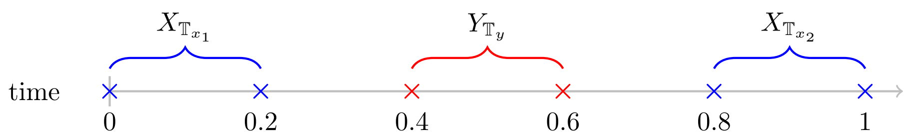


Figure 6: An illustration of Example 4, showing the temporal relationship between two different solar flare variables $X_{\mathbb{T}_{x_1}}$ and $X_{\mathbb{T}_{x_2}}$ (both blue), and power outage $Y_{\mathbb{T}_y}$ (red). Total effect-acyclicity by Definition 5 holds for the causal relationship between $X_{\mathbb{T}_{x_1}}$ and $Y_{\mathbb{T}_y}$, and between $X_{\mathbb{T}_{x_2}}$ and $Y_{\mathbb{T}_y}$, since power outages cannot affect solar flares irrespective of the temporal relationship.

Figure 6 illustrates the example. For the solar flare variable preceding the power outage variable there may be a path $X_{\mathbb{T}_{x_1}} \rightsquigarrow Y_{\mathbb{T}_y}$, but there is no path in either direction between $Y_{\mathbb{T}_y}$ and the solar flare variable $X_{\mathbb{T}_{x_2}}$ succeeding it in time. More generally, no temporal component of power outage causes any temporal component of solar flares, hence total effect-acyclicity holds by Definition 5. Only when there is total effect-acyclicity between two causal variables does a path in one direction rule out paths in the other direction for all temporal components. Note the contrast to the current paradigm of nontemporal DAGs, in which a path $X \rightsquigarrow Y$ categorically excludes any path $Y \rightsquigarrow X$.

The scenario chosen here also illustrates that examples for which the general absence of causal paths in one direction is certain often involve processes for which realistic interventions are difficult to conceive, in this case on the solar flares (for examples of cause-effect pairs see e.g. Mooij et al. 2016).

4 Assessing Acyclicity and Faithfulness

We outline how time-acyclicity (Definition 3) and effect-acyclicity (Definition 4) offer different ways for assessing the validity of the acyclicity assumption, and demonstrate that they provide a natural framework for unrolling cyclic causal relationships in time. Additionally, we show that the explicit treatment of time in our formalism can help uncover faithfulness violations.

4.1 Assessing the Validity of the Acyclicity Assumption

The acyclicity assumption is a defining characteristic of DAGs. For a DAG to be a fitting model of the causal relationships between some variables, the acyclicity assumption must be valid for these variables. The concepts of time-acyclicity and effect-acyclicity can be used for assessing this validity.

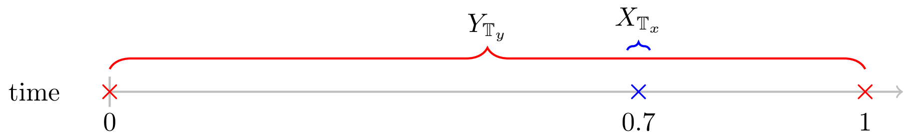
The causality literature often assumes acyclicity directly, rather than any specific reason as to why it may be valid. In this more general approach, it remains open how to operationalize assessing the validity of the acyclicity assumption. For a prominent example, Spirtes et al. (2001), Chapter 5.4 states that their algorithms, including the popular PC algorithm, “eliminate the need for a prior ordering of the variables”. This contrasts with earlier approaches (e.g. Wermuth and Lauritzen 1983), which ensure acyclicity through a prior ordering, typically by time. However, the algorithms by Spirtes et al. (2001) still assume acyclicity, even if it no longer needs to be explicitly ensured by a time order. The assumption of acyclicity could thus be read as the assertion that a time order that prevents cycles *exists* even if it may be *unknown*. At the same time, cyclic causal relationships such as feedback loops are frequently discussed in the context of causal graphical models (see forthcoming Section 5), including in the same work Spirtes et al. (2001), Chapter 12. From Definition 3 it follows that causal cycles require the absence of a time order. If the time order is *unknown*, it may thus be that it *does not exist* at all, meaning that cycles are possible and acyclicity relies on their absence.

The concepts of time-acyclicity (Definition 3) and effect-acyclicity (Definition 4) capture this difference and lead to correspondingly different approaches for assessing the validity of the acyclicity assumption. Assessing time-acyclicity involves examining whether the support of the joint distribution of time points of the variables in question provides a time order that prevents cycles. This is principally an analysis of the measurement procedure. Assessing effect-acyclicity requires additionally examining potential causal cycles where the time order does not prevent them. This necessitates an analysis of the causal relationships between the variables of the underlying stochastic processes. Thus, time-acyclicity and effect-acyclicity provide two concrete ways of assessing the validity of the acyclicity assumption, each relying on different background knowledge. The explicit role of time in our formalism clarifies the practical feasibility of such an assessment by specifying measurement times, thereby highlighting acyclicity as a testable assumption.

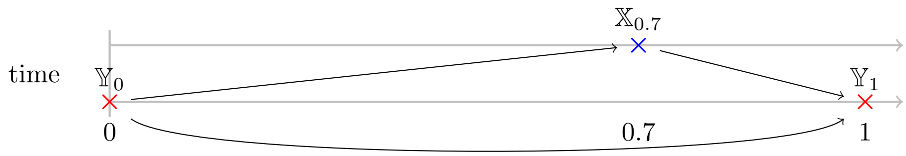
4.2 Unrolling Causal Cycles

It is often suggested that cyclic causal relationships may be acyclic at a sufficiently fine-grained temporal resolution (e.g. Pearl 2009, Chapter 1.4.1; Peters et al. 2017, Chapter 2.3.3). In other words, one could ‘unroll’ cycles in time to obtain a DAG. Such an operation relies on a correspondence between variables and time points and effectively postulates the existence of an atomic DAG, which is precisely what we aim to make explicit in this work. Resolving cycles by unrolling them in time can be understood as addressing a lack of effect-acyclicity (Definition 4) between some variables by splitting them up in time to obtain time-acyclicity (Definition 3) between the so-created variables. Using our formalism, we give a simple example of how a cyclic effect can be unrolled in time to obtain a DAG. We caution, however, that unrolling an aggregate causal variable in time requires access to its temporal components and results in different variables.

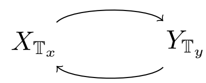
Example 5 (unrolling in time). Consider another variation of our running medical treatment-health example. Let $X_{\mathbb{T}_x}$ be an aggregate of $k_x = 1$ health measurement in $(\mathbb{X})_{t\in\mathcal{T}_x}$, where $\mathcal{T}_x = \{0.7\}$. Let $Y_{\mathbb{T}_y}$ be an aggregate of $k_y = 2$ treatment measurements in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, where $\mathcal{T}_y = \{0, 1\}$. It follows that $\mathbb{T}_x = \mathcal{T}_x$, $\mathbb{T}_y = \mathcal{T}_y$, and $\mathcal{S}_{xy} = \{(\{0.7\}, \{0, 1\})\}$, meaning that measurements are always taken at all time points. Let the aggregation function f_x be the identity, and let f_y take the sum of its inputs. Assume that this setting corresponds to an initial health measurement that informs a treatment, which leads to a final health measurement. Assume further that initial health also affects final health directly. This is to say that $\mathbb{Y}_0$ causes $\mathbb{X}_{0.7}$, $\mathbb{X}_{0.7}$ causes $\mathbb{Y}_1$, and $\mathbb{Y}_0$ causes $\mathbb{Y}_1$ in the atomic DAG.



(a) Timeline and variable aggregation; there is no time-acyclicity by Definition 3.



(b) Atomic causal DAG.

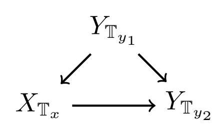


(c) Cyclic causal directed graph.

Figure 7: The timeline, atomic causal DAG, and causal directed graph of Example 5, where the variables $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ have a cyclic causal relationship.

Figure 7 illustrates the example. The temporal aggregation (Figure 7a) of the atomic causal DAG (Figure 7b) gives rise to a causal cycle (Figure 7c). The causal cycle can be resolved if the health variable is split into two parts: one representing health before the treatment, and one after. For health before treatment, let $Y_{\mathbb{T}_{y_1}}$ be an aggregate of $k_{y_1} = 1$ health measurement in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, such that $\mathcal{S}_{y_1} = \{(\{0\})\}$. For health after treatment, let $\mathbb{Y}_{\mathbb{T}_{y_2}}$ be an aggregate of $k_{y_2} = 1$ health measurement in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, such that $\mathcal{S}_{y_2} = \{(\{1\})\}$.

(a) New causal variables with time-acyclicity by Definition 3.



(b) New causal DAG.

Figure 8: Resolving the causal cycle by splitting $Y_{\mathbb{T}_y}$ in time.

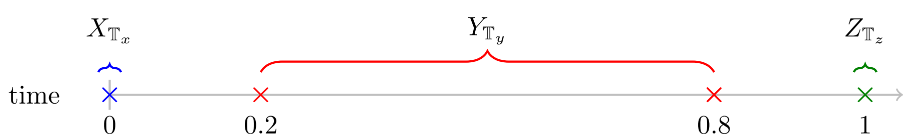
The result can be seen in Figure 8. Splitting $Y_{\mathbb{T}_y}$ into $Y_{\mathbb{T}_{y_1}}$ and $Y_{\mathbb{T}_{y_2}}$, as shown in Figure 8a, transforms the cyclic causal relationship into the confounder-DAG between the new variables shown in Figure 8b. It is well known that the past of a variable is a natural candidate for a confounder (see e.g. Imbens and Rubin 2015, Chapter 1.8). Our formal treatment of time captures this idea in an explicit and systematic fashion.

Note, however, that unrolling a causal cycle requires that measurements at a higher temporal resolution are available. Moreover, the resulting causal DAG features different causal variables that may not be equally suitable for the purpose of analysis as the original ones (see also Hyttinen et al. 2012, Section 1; Mooij et al. 2011, Section 1). Aggregate variables of settings characterized by high-frequency interactions, such as stock prices or gene expressions, may therefore be difficult to unroll into meaningful DAGs.

4.3 Uncovering Faithfulness Violations

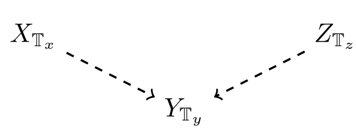
The foundational idea for discovering causal DAGs from data is to test for conditional independencies between a set of variables to find a Bayesian network which may be interpreted as the causal DAG (or an equivalence class containing the DAG) describing the data generating process; we refer to Spirtes et al. (2001) for a detailed treatment. This approach relies upon the causal faithfulness assumption (see e.g. Peters et al. 2017, Definition 6.33). Using our formalism, we give an example that features a Bayesian network edge pointing backward in time to demonstrate that the aggregation of multiple time points may result in faithfulness violations. Faithfulness violations arising from variable groupings are described in Parviainen and Kaski (2017), Section 3 and further analyzed in Wahl et al. (2024), Sections 5 and 6. This type of faithfulness violation cannot be easily dismissed despite the arguments in Pearl (2009), Chapter 2.4, Spirtes et al. (2001), Chapter 3.5.2 and Theorem 3.2, and Meek (1995b).

Example 6 (faithfulness violation). Let $X_{\mathbb{T}_x}$ be an aggregate of $k_x = 1$ measurement of Aspirin intake in $(\mathbb{X}_t)_{t\in\mathcal{T}_x}$, where $\mathcal{T}_x = \{0\}$, meaning that $\mathbb{T}_x = \mathcal{T}_x$. Let $Y_{\mathbb{T}_y}$ be an aggregate of $k_y = 2$ measurements of headache in $(\mathbb{Y}_t)_{t\in\mathcal{T}_y}$, where $\mathcal{T}_y = \{0.2, 0.8\}$, meaning that $\mathbb{T}_y = \mathcal{T}_y$. Let further $Z_{\mathbb{T}_z}$ be an aggregate of $k_z = 1$ measurement of work attendance in $(\mathbb{Z}_t)_{t\in\mathcal{T}_z}$, where $\mathcal{T}_z = \{0\}$, meaning that $\mathbb{T}_z = \mathcal{T}_z$. Let f_x and f_z be the identity, and f_y take the sum of its inputs. Assume an atomic DAG encoding that Aspirin at $t = 0$ cures headache at $t = 0.2$, and that another later headache at $t = 0.8$ causes absence from work at $t = 1$.



(a) Timeline and variable aggregation.

(b) Atomic DAG.



(c) Bayesian network between the aggregate variables.

Figure 9: An illustration of Example 6. The aggregation of the atomic causal DAG leads to a Bayesian network between aggregate variables that contains an edge pointing backward in time.

The example is illustrated in Figure 9. The temporal aggregation (Figure 9a) of the atomic causal DAG (Figure 9b) leads to a pairwise causal dependence between $X_{\mathbb{T}_x}$ and $Y_{\mathbb{T}_y}$ and between $Y_{\mathbb{T}_y}$ and $Z_{\mathbb{T}_z}$, but not between $X_{\mathbb{T}_x}$ and $Z_{\mathbb{T}_z}$. Applying the orientation rules in Meek (1995a) yields the Bayesian network structure shown in Figure 9c, which would be found using for example the PC algorithm for causal discovery (see e.g. Spirtes et al. 2001, Chapter 5.4.2). Notably, the edge $Z_{\mathbb{T}_z} \dashrightarrow Y_{\mathbb{T}_y}$ in the Bayesian network does not respect the temporal precedence of the cause by Definition 1, even though all edges in the atomic causal DAG do.

This can be explained by a faithfulness violation. The aggregate variables form the cluster DAG $X_{\mathbb{T}_x} \to \boxed{Y_{\mathbb{T}_y}} \to Z_{\mathbb{T}_z}$ (Anand et al. 2023, Definition 1). Since $X_{\mathbb{T}_x}$ and $Z_{\mathbb{T}_z}$ are independent, as can be seen in the atomic DAG, but are not d-separated by the empty set in the cluster DAG, there is a faithfulness violation with respect to the cluster DAG. This example highlights the relevance of the criticism in Dawid (2010), Section 4.3 that interpreting edges in independence models as causal means reaching beyond the formal semantics in an act of reification, which requires careful justification.

5 Related Literature

Critical perspectives on causal DAGs. Causal DAGs allow for far-reaching causal insights precisely because they are a restrictive model class. In his methodological critique, Dawid (2010) points out that the strong assumptions underlying the model class require adequate justification. Greenland (2010) argues that causal DAGs rely on null-hypotheses that are implausible in many applications. Aalen et al. (2016) question whether DAGs between measurements taken at discrete time points are useful for understanding underlying continuous causal processes. Notably, the works by Greenland (2010) and Aalen et al. (2016), although critical of DAGs in other ways, do not address the directionality of edges or the acyclicity assumption. Writing from the perspective of the applied sciences, they naturally assume that the causal variables refer to measurements taken at certain times, giving rise to a time order. This is in contrast with the idea of nontemporal 'equilibrium' variables predominant in the causal DAG literature (see e.g. Pearl 2009, Chapter 2; Spirtes et al. 2010, Section 5; Peters et al. 2017, Remark 6.7). The problem with the nontemporal approach, namely that generic variables can refer to items of varying temporal positions, is noted in Spohn (2010), Section 3.6. We adopt the applied sciences' pragmatic view that measurements are necessarily time-specific, and take the conceptual problem of generic variables to its logical conclusion: causal relationships are relative with respect to time. This motivates our formalization of the time of causal variables, highlights acyclicity as a testable assumption in need of justification, and clarifies the type of causal relationships encoded by causal DAGs.

Cyclic causal graphs. Since many systems of interest may be characterized by cyclic causal relationships (e.g. feedback loops), cyclic causal graphical models are an active area of research. Although most of the central properties of DAGs do not hold in general if there are cycles, many of them may hold for certain subclasses of cyclic graphs (see Bongers et al. 2021). The study of cyclic systems is often motivated by the fact that, even if causal relationships were acyclic at a fine-grained temporal resolution, they may be cyclic at the level of the variables of interest, (see e.g. Hyttinen et al. 2012; Bongers et al. 2021, who both give the example of economic supply and demand). In this respect, our work can be seen as a complement to the literature on cyclic causal graphs. Characterizing the role of time for the assumption of acyclicity in causal DAGs provides guidance as to what settings can be analyzed using causal DAGs, and what settings may require other approaches such as those involving cyclic causal graphs.

Time-series DAGs. Causal modeling of time series, following Granger (1969), considers causal relationships in dynamical systems that are sampled at regular intervals. In contrast to the setting of mutually independent samples usually modeled by causal graphs (Spirtes et al. 2001; Pearl 2009), this setting is characterized by time dependency between samples. Nonetheless, causal graphs have been applied to time series data, in particular for the task of causal discovery (e.g. Peters et al. 2013; Malinsky and Spirtes 2018; Runge 2018). The adaptation of causal graphical models to this context necessitates an explicit treatment of time, which has lead to the adoption of dynamic Bayesian networks (Dean and Kanazawa 1989; Murphy 2002). Peters et al. (2013) introduce the distinction between a 'full-time' graph between all variables at all time points, and a 'summary' graph, which combines the time points into a single node for each variable. The use of a fixed maximum time-lag allows extracting a finite repeating 'window' graph from the otherwise infinite full-time graph. For a detailed treatment of time-series causality, we refer to Assaad (2021). The nature of time series imposes a temporal order between variables at the level of the window graph, which is assumed to be a DAG. The summary graph, by contrast, does not contain temporal information and may be cyclic. Our approach can be seen as a bridge between nontemporal and time-series causality. We highlight that, even if samples are mutually independent and no matter the level of aggregation, causal variables always have a time dimension which is crucial for interpreting causal relationships and for assessing the validity of the acyclicity assumption. Thus, just as in the time-series setting, DAGs are generally more readily admissible for temporally separated variables than for aggregate variables of overlapping time intervals.

Continuous-time causal models. Hansen and Sokol (2014) propose a causal interpretation of stochastic differential equations (SDEs). Specifically, they show convergence in post-interventional distribution of the structural causal model given by Euler discretization to the solution of the SDE. An SDE can thus be interpreted as a continuous-time generalization of a full-time causal DAG. Lorch et al. (2024) posit the fact that SDEs do not assume acyclicity between the aggregate variables as an advantage over conventional causal DAGs. Since SDEs can be understood as implicitly unrolling any cyclic relationships in the summary graph into a full-time DAG, as shown by Hansen and Sokol (2014), the perspective of potentially cyclic graphs between aggregate variables and acyclic graphs between time-unrolled variables unites continuous-time causal models and time-series DAGs. Our work shows that the same perspective can be applied to conventional causal DAGs. Once the time dimension of variables is taken into account, it is clear that defining temporally separated variables is a way of meeting the acyclicity assumption. Where such a separation is not possible, the assumption of acyclicity is indeed more restrictive and requires additional justification.

Time in the potential outcomes framework. In the potential outcomes framework (Splawa-Neyman 1932; Rubin 1974) time is integral to the definition of treatment and outcome, the cause and effect of interest. Treatment and outcome correspond to time points t_1 and t_2 such that $t_1 < t_2$, and the causal effect is defined as the difference in the potential outcomes under treatment and control for these time points. The impossibility of returning from t_2 to t_1 for the same subject and exploring another potential outcome is precisely what is referred to as the fundamental problem of causal inference (Holland 1986). The essential role of time is also reflected in the definition of causal variables as time-specific units, and in the distinction between pre- and post-treatment variables (see e.g. Imbens and Rubin 2015, Chapter 1). Nonetheless, the potential outcomes framework does not generally employ an explicit notion of time for all variables, other than where it is unavoidable, such as in the case of panel data (see e.g. Angrist and Pischke 2009). This reflects the fact that the potential outcomes framework is conceived around a particular causal effect of interest and does not make explicit causal statements about every pair of variables involved. Causal DAGs, by contrast, model the causal relationships between all variables, which offers the advantage of providing clear criteria for selecting control variables (Pearl and Mackenzie 2018, Chapter 4). However, for all their promise, causal DAGs have so far neglected to formally incorporate an essential aspect of causation that is built into the core of the potential outcomes framework and may explain its enduring appeal over DAGs in many applied fields: the notion of time. Our contribution addresses this limitation by incorporating time into the definition of variables in causal DAGs.

6 Time in Causal DAGs: Implications

Causal dicovery and representations. Causal discovery aims to infer causal graphs, for the most part DAGs or their equivalence classes, from data (for an overview see Kitson et al. 2023). Assume for a moment that one were sure that the acyclicity assumption holds, and consider the task of causal DAG discovery without background knowledge. If acyclicity were due to time-acyclicity, one would be sure *that* there is a time order between the variables, but not *what* this order is. If acyclicity were due to effect-acyclicity, one would understand enough about the causal relationships to rule out cycles, but know nothing yet about their directions. In either case, the premise of causal DAG discovery in a vacuum seems tenuous. It would be a curious setting where one could be certain that there is a causal DAG, without already knowing rather a lot about it. Causal DAG discovery without knowing whether the acyclicity assumption holds presents evident challenges. Put like this, the absence of notable novel causal discoveries so far may seem less surprising. We consider the incorporation of time into causal DAGs a necessary and promising avenue for advancing the applicability of causal discovery algorithms from synthetic data to real-world use cases (Gentzel et al. 2019; Reisach et al. 2021; Reisach et al. 2023; Brouillard et al. 2024). Discovering causal structures also includes the search for suitable causal representations to form the basis for such structures (Schölkopf et al. 2021). In this context, our contribution can be understood to add the time dimension as a crucial aspect of causal representations.

Causal DAGs for effect estimation and prediction. The temporal relationship between variables may already provide useful information for effect estimation and prediction, even without a causal DAG. For the purpose of effect estimation, the time order can yield natural candidates for control variables, such as the past of the outcome variable, and rule out some others, e.g. if they are post-outcome. This may explain why the potential outcomes framework (Splawa-Neyman 1932; Rubin 1974) enjoys great popularity for causal inference in the applied sciences (see e.g. Imbens and Rubin 2015), seemingly without a need for DAGs to select control variables. Nonetheless, DAGs provide a principled way of selecting control variables beyond what can be derived from the time order, and incorporating time into causal DAGs may outline this advantage more clearly. Concerning the task of robust prediction, taking the time order into account raises some questions regarding the promises causal DAGs may seem to offer for machine learning (Schölkopf 2022). For example, it would be hard to achieve causal robustness, invariance, etc. if the time points are not the same across regimes, or the causal relationships between aggregate variables do not form a DAG to begin with. This could explain why the practical usefulness of causally robust machine learning remains to be demonstrated (Nastl and Hardt 2024).

Toward applicable causal DAGs. Finally and perhaps most importantly, taking into account the role of time means tackling the question of the applicability of causal DAGs to real-world data in earnest. Before the question "what is the right DAG?", comes the question "is there a DAG?". Time-acyclicity poses a severe restriction on the measurement process of variables, and effect-acyclicity on the causal mechanisms between them. Many variables of interest may therefore not be amenable to causal analysis using DAGs, and what causal DAGs there are, are of a more specific nature allowing for less sweeping conclusions. In this light, the crucial step in applying causal DAGs consists in defining meaningful variables that allow for causal DAGs, rather than treating a set of variables as given and reasoning about *the* DAG between them. An explicit notion of time facilitates finding such variables and ensuring the acyclicity assumption is met, for example by establishing measurement protocols that satisfy time-acyclicity. Incorporating time into causal DAGs therefore offers the potential for a more pragmatic approach to causal inference in line with the practices of applied research.

7 Conclusion: The Case for Time

Real-world quantities and their measurements exist and change in time. If causes are to precede effects, it follows that causal relationships are relative with respect to time. For this reason, we argue that it is time (pun intended) to break with the tradition of nontemporal causal models and incorporate time into the definition of variables in causal DAGs by default. This perspective reveals that causal DAGs do not relate general concepts but rather time-specific instantiations, which is essential for their interpretation and reconciles an important difference between causal analysis based on DAGs versus on potential outcomes. Instantiations that overlap in time may contain causal cycles. In order to establish the validity of the acyclicity assumption, one must show that the variable measurements correspond to a temporal order that prevents causal cycles (time-acyclicity), or that there are no causal cycles where the time order does not prevent them (effect-acyclicity). Incorporating time into causal DAGs is therefore essential for evaluating their applicability and interpretation as causal models. Our formal treatment of time allows for doing so in a principled fashion and opens many questions for future research. For example, it will be interesting to examine the change of causal effects with time distance and to what extent some time points may be representative of others. Moreover, further research is required to analyze the implications in terms of faithfulness violations, and to apply our notion of time to other classes of causal graphical models. Although our contribution shows the scope of causal DAGs to be more specific than it may otherwise appear, it also facilitates their discovery, clarifies their interpretation, and leaves them a less metaphysical and all the more practical tool for causal analysis.

Acknowledgements

We thank Niels Richard Hansen, Emilie Devijver, and Pouya Babakhani for helpful discussions. AGR received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945332 and travel support under No 951847, as well as mobility funding from Université Paris Cité via the 2024 BDMI program.

References

[1] Odd O. Aalen, Kjetil Røysland, Jon M. Gran, Roger Kouyos, and Tanja Lange. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. In: *Statistical Methods in Medical Research* 25.5 (2016), pp. 2294–2314 (cit. on p. 15).

[2] Tara V. Anand, Adele H. Ribeiro, Jin Tian, and Elias Bareinboim. Causal Effect Identification in Cluster DAGs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 10. 2023, pp. 12172–12179 (cit. on pp. 6, 15).

[3] Joshua D. Angrist and Jörn-Steffen Pischke. Mostly harmless econometrics: An empiricist's companion. Princeton University Press, 2009 (cit. on p. 17).

[4] Charles Assaad. Causal Discovery between time series. Theses. Université Grenoble Alpes, July 2021 (cit. on pp. 4, 16).

[5] Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M. Mooij. Foundations of Structural Causal Models with Cycles and Latent Variables. In: *The Annals of Statistics* 49.5 (2021), pp. 2885–2915 (cit. on p. 15).

[6] Philippe Brouillard, Chandler Squires, Jonas Wahl, Konrad P. Kording, Karen Sachs, Alexandre Drouin, and Dhanya Sridhar. The Landscape of Causal Discovery Data: Grounding Causal Discovery in Real-World Applications. In: *arXiv preprint* (2024) (cit. on pp. 9, 17).

[7] Philip A. Dawid. Beware of the DAG! In: *Causality: Objectives and Assessment*. PMLR. 2010, pp. 59–86 (cit. on pp. 3, 15).

[8] Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. In: *Computational intelligence* 5.2 (1989), pp. 142–150 (cit. on p. 16).

[9] Amanda Gentzel, Dan Garant, and David Jensen. The Case for Evaluating Causal Models Using Interventional Measures and Empirical Data. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 32. 2019 (cit. on p. 17).

[10] Clive W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. In: *Econometrica* 37.3 (Aug. 1969), pp. 424–438 (cit. on pp. 2, 16).

[11] Sander Greenland. Overthrowing the Tyranny of Null Hypotheses Hidden in Causal Diagrams. In: *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (2010), pp. 365–382 (cit. on p. 15).

[12] Niels Hansen and Alexander Sokol. Causal interpretation of stochastic differential equations. In: *Electronic Journal of Probability* 19 (Jan. 2014), pp. 1–24 (cit. on p. 16).

[13] Austin B. Hill. The Environment and Disease: Association or Causation? In: *Proceedings of the Royal Society of Medicine* (1965), pp. 295–300 (cit. on p. 2).

[14] Paul W. Holland. Statistics and Causal Inference. In: *Journal of the American statistical Association* 81.396 (1986), pp. 945–960 (cit. on p. 16).

[15] Biwei Huang, Kun Zhang, and Bernhard Schölkopf. Identification of Time-Dependent Causal Model: A Gaussian Process Treatment. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015 (cit. on p. 3).

[16] Antti Hyttinen, Frederick Eberhardt, and Patrik O. Hoyer. Learning Linear Cyclic Causal Models with Latent Variables. In: *Journal of Machine Learning Research* 13.109 (2012), pp. 3387–3439 (cit. on pp. 13, 15).

[17] Guido W. Imbens and Donald B. Rubin. Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge University Press, 2015 (cit. on pp. 2, 4, 13, 17).

[18] Neville K. Kitson, Anthony C. Constantinou, Zhigao Guo, Yang Liu, and Kiattikun Chobtham. A survey of Bayesian Network structure learning. In: *Artificial Intelligence Review* (2023), pp. 1–94 (cit. on p. 17).

[19] Pierre Simon de Laplace. Essai philosophique sur les probabilités. Courcier, 1814, link to the 5th edition from 1825 (cit. on p. 4).

[20] Hao-Chih Lee, Matteo Danieletto, Riccardo Miotto, Sarah T. Cherng, and Joel T. Dudley. Scaling structural learning with NO-BEARS to infer causal transcriptome networks. In: *Pacific Symposium on Biocomputing 2020*. World Scientific. 2019, pp. 391–402 (cit. on p. 2).

[21] Lars Lorch, Andreas Krause, and Bernhard Schölkopf. Causal Modeling with Stationary Diffusions. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2024, pp. 1927–1935 (cit. on p. 16).

[22] Daniel Malinsky and Peter Spirtes. Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding. In: *Proceedings of 2018 ACM SIGKDD workshop on causal discovery*. PMLR. 2018, pp. 23–47 (cit. on p. 16).

[23] Christopher Meek. Causal inference and causal explanation with background knowledge. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. UAI'95. Morgan Kaufmann Publishers Inc., 1995, pp. 403–410 (cit. on p. 14).

[24] Christopher Meek. Strong Completeness And Faithfulness In Bayesian Networks. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. UAI'95. Morgan Kaufmann Publishers, 1995, pp. 411–418 (cit. on p. 14).

[25] Joris M. Mooij, Dominik Janzing, Tom Heskes, and Bernhard Schölkopf. On Causal Discovery with Cyclic Additive Noise Models. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 24. 2011 (cit. on p. 13).

[26] Joris M. Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing Cause from Effect Using Observational Data: Methods and Benchmarks. In: *Journal of Machine Learning Research* 17.32 (2016), pp. 1–102 (cit. on pp. 3, 11).

[27] Kevin Patrick Murphy. Dynamic bayesian networks: representation, inference and learning. University of California, Berkeley, 2002 (cit. on p. 16).

[28] Vivian Y. Nastl and Moritz Hardt. Do causal predictors generalize better to new domains? In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 37. 2024 (cit. on p. 18).

[29] Pekka Parviainen and Samuel Kaski. Learning structures of Bayesian networks for variable groups. In: *International Journal of Approximate Reasoning* 88 (2017), pp. 110–127 (cit. on p. 14).

[30] Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000 (cit. on p. 2).

[31] Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2009 (cit. on pp. 2–4, 12, 14–16).

[32] Judea Pearl and Dana Mackenzie. The Book of Why: The New Science of Cause and Effect. Basic Books, 2018 (cit. on pp. 4, 17).

[33] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Causal Inference on Time Series using Restricted Structural Equation Models. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 26. 2013 (cit. on p. 16).

[34] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press, 2017 (cit. on pp. 2–5, 12, 14, 15).

[35] Joseph Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. In: *International Journal of Data Science and Analytics* 3 (2017), pp. 121–129 (cit. on p. 2).

[36] Alexander G. Reisach, Christof Seiler, and Sebastian Weichwald. Beware of the Simulated DAG! Causal Discovery Benchmarks May Be Easy to Game. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 34. 2021 (cit. on p. 17).

[37] Alexander G. Reisach, Myriam Tami, Christof Seiler, Antoine Chambaz, and Sebastian Weichwald. A Scale-Invariant Sorting Criterion to Find a Causal Order in Additive Noise Models. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 36. 2023 (cit. on p. 17).

[38] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. In: *Journal of Educational Psychology* 66.5 (1974), 688–701, link to the earlier open access 1972 ETS Technichal Report (cit. on pp. 16, 17).

[39] Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7 (2018) (cit. on p. 16).

[40] Bernhard Schölkopf. Causality for Machine Learning. In: *Probabilistic and Causal Inference: The Works of Judea Pearl*. 2022, pp. 765–804 (cit. on p. 17).

[41] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan R. Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward Causal Representation Learning. In: *Proceedings of the IEEE* 109.5 (2021), pp. 612–634 (cit. on p. 17).

[42] Yoav Shoham. Nonmonotonic reasoning and causation. In: *Cognitive Science* 14.2 (1990), pp. 213–252 (cit. on p. 2).

[43] Ilya Shpitser and Judea Pearl. Complete Identification Methods for the Causal Hierarchy. In: *Journal of Machine Learning Research* 9 (2008), pp. 1941–1979 (cit. on p. 4).

[44] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. In: *Lecture Notes in Statistics* 81 (1993) (cit. on p. 2).

[45] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. MIT Press, 2001 (cit. on pp. 2, 11, 12, 14–16).

[46] Peter Spirtes, Clark Glymour, Richard Scheines, and Robert Tillman. Automated Search for Causal Relations: Theory and Practice. In: *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (2010), pp. 467–502 (cit. on p. 15).

[47] Jerzy Splawa-Neyman. On the Application of Probability Theory to Agricultural Experiments. In: *Statistical Science* 5 (1932), pp. 465–480 (cit. on pp. 16, 17).

[48] Wolfgang Spohn. The Structural Model and the Ranking Theoretic Approach to Causation: A Comparison. In: *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (2010), pp. 507–522 (cit. on p. 15).

[49] Patrick Suppes. A Probabilistic Theory of Causality. North-Holland Publishing Company, 1970 (cit. on p. 2).

[50] Jonas Wahl, Urmi Ninad, and Jakob Runge. Foundations of causal discovery on groups of variables. In: *Journal of Causal Inference* 12.1 (2024), p. 20230041 (cit. on p. 14).

[51] Nanny Wermuth and Steffen L. Lauritzen. Graphical and Recursive Models for Contingency Tables. English. In: *Biometrika* 70.3 (1983), pp. 537–552 (cit. on p. 12).

[52] Sewall Wright. The Method of Path Coefficients. In: *The Annals of Mathematical Statistics* 5.3 (1934), pp. 161–215 (cit. on p. 2).