



HAL
open science

Causal Inference Theory with Information Dependency Models

Benjamin Heymann, Michel de Lara, Jean-Philippe Chancelier

► **To cite this version:**

Benjamin Heymann, Michel de Lara, Jean-Philippe Chancelier. Causal Inference Theory with Information Dependency Models. 2021. hal-03315815v2

HAL Id: hal-03315815

<https://hal.science/hal-03315815v2>

Preprint submitted on 7 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Causal Inference Theory with Information Dependency Models

Benjamin Heymann*, Michel De Lara†, Jean-Philippe Chancelier†

August 7, 2021

Abstract

Inferring the potential consequences of an unobserved event is a fundamental scientific question. To this end, Pearl’s celebrated do-calculus provides a set of inference rules to derive an interventional probability from an observational one. In this framework, the primitive causal relations are encoded as functional dependencies in a Structural Causal Model (SCM), which are generally mapped into a Directed Acyclic Graph (DAG) in the absence of cycles. In this paper, by contrast, we capture causality without reference to graphs or functional dependencies, but with information fields and Witsenhausen’s intrinsic model. The three rules of do-calculus reduce to a unique sufficient condition for conditional independence, the topological separation, which presents interesting theoretical and practical advantages over the d-separation. With this unique rule, we can deal with systems that cannot be represented with DAGs, for instance systems with cycles and/or ‘spurious’ edges. We treat an example that cannot be handled — to the extent of our knowledge — with the tools of the current literature. We also explain why, in the presence of cycles, the theory of causal inference might require different tools, depending on whether the random variables are discrete or continuous.

1 Introduction

As the world shifts toward more and more data-driven decision-making, causal inference is taking more space in applied sciences, statistics and machine learning. This is because it allows for better, more robust decision-making, and provides a way to interpret the data that goes beyond correlation [17]. For instance, causal inference provides a language to describe and solve Simpson’s paradox, which embodies the “correlation is not causation” principle as can be found in any “Statistics 101” basic course. The main concern in causal inference is to compute post-intervention probability distributions from observational data. For this

*Criteo AI Lab, Paris, France

†CERMICS, Ecole des Ponts, Marne-la-Vallée, France

purpose, graphical models are practical because they allow representing assumptions easily and benefit from an extensive scientific literature.

In his seminal work [15], Pearl builds on graphical models [8] to introduce the so-called do-calculus. Causal graphical models move the focus from joint probability distributions to functional dependencies thanks to the Structural Causal Model (SCM) framework. Several extensions to this do-calculus have been proposed recently [25, 14, 23, 7]. Pearl’s seminal paper supposes a Directed Acyclic Graph (DAG) structure.

In this paper, we bring a new, complementary view to the causal reasoning toolbox by leveraging the concept of information fields and Witsenhausen’s intrinsic model. The framework we introduce is general, unifying, and may be used to study causal inference in both recursive and nonrecursive systems [13] (i.e. with and without cycles). It allows for spurious edges, and simplifies the statement of Pearl’s three rules of do-calculus.

DAGs modeling does not rely directly on random variables but on joint probability distributions (see [16, footnote 3] or [18, Appendix A]). By contrast, our approach requires going back to the classical primitives of probabilistic models: sample sets, σ -fields, measurable maps and random variables. We exploit the generally overlooked expressiveness of this underlying structure. The cost for this conceptual generalization is a bit of abstraction: in what we propose, the structure is implicit, and there are no arrows.

This paper, however, has been written so that the main messages can be understood with the usual graphical concepts used in the field of causal inference: the notion of topological separation is explained for the specific case of DAGs; Theorem 25 and Examples 10, 11 and 13 should be readable without the concept of information field. In addition, this paper was written in parallel to two other papers [6, 9]; the three of them aim at providing another perspective on conditional independence and do-calculus.

Related work and contributions. We extend the causal modeling toolbox thanks to two notions: information fields and topological separation. These two notions rely on the foundational work produced by Witsenhausen in the seventies [26]. The concept of information field extends the expressiveness of the Structural Causal Model, and allows for instance to naturally encode context specific independence [23]. In the companion papers [6, 9], we show an equivalence between Pearl’s d-separation and a new notion that we introduce, the conditional topological separation. The topological separation is practical because it just requires to check that two sets are disjoint (see Examples 2). By contrast, the d-separation requires to check that *all* the paths that connect two variables are blocked. Moreover, as its name suggests, the topological separation has a theoretical interpretation. Specifically, the topological separation allows us to go beyond DAGs and even graphical models.

Our main results are (i) Theorem 25, which is a generalization of do-calculus that can be applied in particular to nonrecursive systems [4] and which subsumes several recent results, and (ii) Lemma 20 which provides insight into the machinery behind Theorem 25. We pinpoint the novelty of our approach with Example 13, a system with cycles where our framework identifies a probabilistic independence that the framework developed in [12] (for cycles) does not. We explain in Sect. 5.4 that the differences between the framework

developped in [12] and ours comes from a fundamental difference whether in the discrete or in the continuous setting regarding random variables.

The paper is organized in two parts as follows. First, we provide what we think will be of interest for application minded researchers in Sect. 2 and 3. Sect. 2 introduces the notion of Information Dependency Model, which is another way of looking at systems that can be represented with SCMs. We then presents our main results in Sect. 3: we restate Pearl’s do-calculus theorem in terms of topological separation. Second, we present the theoretical foundation of those results in Sect. 4 and 5. Sect. 4 present Witsenhausen’s intrinsic model upon which we build our contributions. We provide the proofs in Sect. 5.

2 Definition of Information Dependency Models

In §2.1, we provide background on σ -fields and introduction the Information Dependency Models. Then, in §2.2, we define conditional precedence.

2.1 Information fields and Information Dependency Models

We start with a few reminders from measure (and probability) theory. A σ -field (henceforth sometimes referred to as *field*) over a set \mathbb{D} is a subset $\mathcal{D} \subset 2^{\mathbb{D}}$, containing \mathbb{D} , and which is stable under complementation and under countable union. The couple $(\mathbb{D}, \mathcal{D})$ is called a *measurable space*. The trivial σ -field over the set \mathbb{D} is $\{\emptyset, \mathbb{D}\}$. The complete σ -field over the set \mathbb{D} is $2^{\mathbb{D}}$. When $\mathcal{D}' \subset \mathcal{D}$ are two σ -fields over the set \mathbb{D} , we say that \mathcal{D}' is a *subfield* of \mathcal{D} . If \mathcal{D} is a σ -field over the set \mathbb{D} and if $\mathbb{D}' \subset \mathbb{D}$, then $\mathcal{D} \cap \mathbb{D}' = \{D \cap \mathbb{D}' \mid D \in \mathcal{D}\}$ is a σ -field over the set \mathbb{D}' , called the *trace subfield* of \mathcal{D} over \mathbb{D}' . If $(\mathbb{D}_i, \mathcal{D}_i)$, $i = 1, 2$ are two measurable spaces, we denote by $\mathcal{D}_1 \otimes \mathcal{D}_2$ the *product σ -field* on $\mathbb{D}_1 \times \mathbb{D}_2$ generated by the rectangles $\{D_1 \times D_2 \mid D_i \in \mathcal{D}_i, i = 1, 2\}$. More generally, if $\{(\mathbb{D}_s, \mathcal{D}_s)\}_{s \in \mathbb{S}}$ is a family of measurable spaces, we denote by $\bigotimes_{s \in \mathbb{S}} \mathcal{D}_s$ the *product σ -field* on $\prod_{s \in \mathbb{S}} \mathbb{D}_s$ generated by the cylinders. Let (Ω, \mathcal{F}) and $(\mathbb{U}, \mathcal{U})$ be two measurable spaces, probability theory defines a *random variable* as a measurable mapping from (Ω, \mathcal{F}) to $(\mathbb{U}, \mathcal{U})$, that is, a mapping $\lambda : \Omega \rightarrow \mathbb{U}$ satisfying $\lambda^{-1}(\mathcal{U}) \subset \mathcal{F}$. When equipped with a probability, \mathbb{P} , a measurable space (Ω, \mathcal{F}) is called a probability space and is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1.1 Structural Causal Models (informal definition)

Thus equipped, we now discuss the standard way to model causal hypotheses using *Structural Causal Models (SCMs)* [18].

Let \mathbb{A} be a set and, for each $a \in \mathbb{A}$, a given probability space $(\Omega_a, \mathcal{F}_a, \mathbb{P}_a)$. We consider the product probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \prod_{a \in \mathbb{A}} \Omega_a$, $\mathcal{F} = \bigotimes_{a \in \mathbb{A}} \mathcal{F}_a$, and $\mathbb{P} = \bigotimes_{a \in \mathbb{A}} \mathbb{P}_a$. Let $\{(\mathbb{U}_a, \mathcal{U}_a)\}_{a \in \mathbb{A}}$ be a family of measurable spaces.

An SCM consists of a family $(\lambda_a)_{a \in \mathbb{A}}$ of mappings (or *assignments*), where each λ_a has codomain \mathbb{U}_a , alongside with a parental mapping $P : \mathbb{A} \rightarrow 2^{\mathbb{A}}$, and of a family of random

variables $(\mathbf{U}_a)_{a \in \mathbb{A}}$, all defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and such that each \mathbf{U}_a has codomain \mathbb{U}_a , with the property that

$$\mathbf{U}_a(\omega) = \lambda_a(\omega_a, \mathbf{U}_{P(a)}(\omega)) , \quad \forall \omega \in \Omega , \quad \forall a \in \mathbb{A} , \quad (1)$$

where ω_a is the projection of ω on Ω_a .

To get the *graphical representation* of a SCM — as a subgraph of the graph $(\mathbb{A}, \mathbb{A} \times \mathbb{A})$ — we draw an arrow $a \rightarrow b$ whenever $a \in P(b)$. Usually, the graphical representation is assumed to be a DAG, which means that the parental mapping induces a partial order on the set \mathbb{A} . We will not need this assumption here. Sufficient conditions to obtain causal properties, relying only on the graphical representation (which is uniquely defined by the parental mapping P), have been developed by many authors. These conditions take their importance from the fact that they short-circuit reasoning on the assignment mappings. For a given applied problem, the SCM is derived from expert knowledge, assumptions and data analysis methods. The SCM is a central tool in causal analysis but its graphical representation does not naturally account for situations such as Context Specific Independence (see [23], and Example 7), where some edges are spurious.

2.1.2 Information Dependency Models (first informal definition)

From Equation (1), the set of arguments of the assignment mapping λ_a depends on a in the formalism of the SCM, (remember that λ_a is the assignment function of \mathbf{U}_a for some $a \in \mathbb{A}$). By contrast, in the Information Dependency Model formulation, the assignment mappings have a common domain, that we call the *configuration space*, which is the product space¹ $\mathbb{H} = \Omega \times \prod_{a \in \mathbb{A}} \mathbb{U}_a$. The *configuration field* $\mathcal{H} = \mathcal{F} \otimes \bigotimes_{a \in \mathbb{A}} \mathcal{U}_a$ is a σ -field over \mathbb{H} . We then extend the definition of SCM thanks to the following observation: we can express the SCM in §2.1.1 by saying that λ_a is a map from \mathbb{H} to \mathbb{U}_a ($\lambda_a : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_a, \mathcal{U}_a)$) while imposing that λ_a “only depends on $\mathbf{U}_{P(a)}$ and ω_a ”. It is standard (see [10, Chap. 1 p. 18]) in probability theory that such property is — under mild assumptions — equivalent to a so-called *measurability constraint* on the random variable \mathbf{U}_a . Hence, the informal definition (1) of a SCM can be restated as

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{F}_a \otimes \bigotimes_{b \neq a} \{\emptyset, \Omega_b\} \otimes \bigotimes_{b \in P(a)} \mathcal{U}_b \otimes \bigotimes_{b \notin P(a)} \{\emptyset, \mathbb{U}_b\} , \quad (2)$$

or, with a slight abuse of notations that we will sometimes use throughout this presentation²

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{F}_a \otimes \bigotimes_{b \in P(a)} \mathcal{U}_b . \quad (3)$$

Informally, an information field is anything one may want to see on the right-hand side of Equation (3). For instance, consider the case where $\mathbb{A} = \{a, b, c\}$ and suppose that all fields contain the singletons. If $\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{F}_a \otimes \{\emptyset, \Omega_b\} \otimes \{\emptyset, \Omega_c\} \otimes \{\emptyset, \mathbb{U}_a\} \otimes \{\emptyset, \mathbb{U}_b\} \otimes$

¹Also called *hybrid space* [26], hence the \mathbb{H} notation.

²We omit the trivial fields in the product on the right-hand side of Equation (3).

$\{\emptyset, \mathcal{U}_c\}$, that we abusively write $\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{F}_a$, this means that $\lambda_a(\omega_a, \omega_b, \omega_c, u_a, u_b, u_c) = \lambda_a(\omega_a, \omega_b, \omega_c, u_a, u_b, u_c)$ only depends on ω_a , that is, only depends on its own “source of uncertainty” (the field \mathcal{F}_a). If (abusively) $\lambda_b^{-1}(\mathcal{U}_b) \subset \mathcal{F}_a \otimes \mathcal{U}_c$, this means that $\lambda_b(\omega_a, \omega_b, \omega_c, u_a, u_b, u_c) = \lambda_b(\omega_a, \omega_b, \omega_c, u_a, u_b, u_c)$ only depends on (ω_a, u_c) , that is, only depends on the uncertainty ω_a (the field \mathcal{F}_a) and on the variable u_c (the field \mathcal{U}_c). More complex examples will be given later.

After having discussed how SCMs can be interpreted with the help of information fields, we propose the name Information Dependency Model for their extension.

Definition 1 (Information Dependency Model) *An Information Dependency Model (IDM) is a collection $(\mathcal{J}_a)_{a \in \mathbb{A}}$ of subfields of \mathcal{H} such that, for any $a \in \mathbb{A}$, $\mathcal{J}_a \subset \mathcal{F}_a \otimes \bigotimes_{b \in \mathbb{A}} \mathcal{U}_b$. The subfield \mathcal{J}_a is called the information field of a .*

The SCM defining property (1) is now expressed in term of the *measurability property*

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{J}_a, \quad \forall a \in \mathbb{A}. \quad (4)$$

Property (4) expresses, in a very general way, that the random variable \mathbf{U}_a may only depend upon the available information \mathcal{J}_a . It is a generalization of the notion of nonanticipativity constraint, or of adapted process with respect to a filtration, in stochastic control. For a given applied problem, like for the SCM, the IDM can be derived from expert knowledge, assumptions and data analysis methods.

Remark 2 *Any SCM can be mapped into an IDM as we obtain from Equation (1) that, for all $a \in \mathbb{A}$, we have that $\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{J}_a$ with $\mathcal{J}_a = \mathcal{F}_a \otimes \bigotimes_{b \in P(a)} \mathcal{U}_b \subset \mathcal{F}_a \otimes \bigotimes_{b \in \mathbb{A}} \mathcal{U}_b$.*

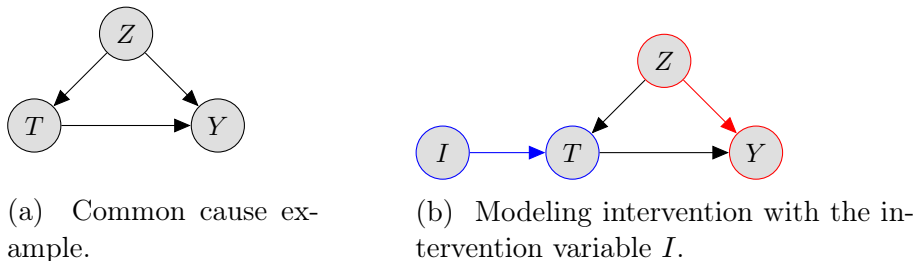


Figure 1: Common cause (Example 3).

Example 3 (Common cause) *First, to better understand how DAGs, and more generally SCMs, can be modeled with information fields, we provide a detailed instance for a set of random variables that can be represented by the DAG in Figure 1. Such an effort is not required in practice, because the measurability properties are fully specified by the DAG for*

such a simple instance. Let $\mathbb{A} = \{Z, T, Y\}$. To simplify the exposition, we suppose that the values of each of the three random variables represented on the DAG belong to $\{0, 1\}$. Then, $\mathbb{U}_Z = \mathbb{U}_T = \mathbb{U}_Y = \{0, 1\}$, each equipped with the complete field $\mathcal{U}_Z = \mathcal{U}_T = \mathcal{U}_Y = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. We take $\Omega = \{0, 1\}^3$ as Nature set, equipped with the complete field $\mathcal{F} = 2^\Omega$ made of all subsets of Ω . We write $\Omega = \Omega_Z \times \Omega_T \times \Omega_Y$, where $\Omega_Z = \Omega_T = \Omega_Y = \{0, 1\}$, and $\mathcal{F} = \mathcal{F}_Z \otimes \mathcal{F}_T \otimes \mathcal{F}_Y$, where $\mathcal{F}_Z = \mathcal{F}_T = \mathcal{F}_Y = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. To represent, for instance, the arrows pointing to Y in the DAG in Figure 1a (as well as implicit assumptions about information on Nature), we require that the information field \mathcal{I}_Y satisfies $\mathcal{I}_Y \subset \{\emptyset, \Omega_Z\} \otimes \{\emptyset, \Omega_T\} \otimes \mathcal{F}_Y \otimes \mathcal{U}_Z \otimes \mathcal{U}_T \otimes \{\emptyset, \mathbb{U}_Y\}$. This relation expresses that the information of Y depends at most on its own “source of uncertainty” (the field \mathcal{F}_Y) and on the decisions of both Z and T (the field $\mathcal{U}_Z \otimes \mathcal{U}_T$). Again, the effort of describing explicitly the information field is not required in the case of DAGs, because the mapping from DAGs to IDMs is trivial. On the other hand the IDM allows to express more sophisticated hypotheses.

2.2 Conditional precedence (informal definition)

We now exploit the flexibility of the concept of information field to extend the definition of precedence. For any subset $B \subset \mathbb{A}$, let $\mathcal{H}_B = \mathcal{F} \otimes \bigotimes_{b \in B} \mathcal{U}_b \subset \mathcal{H}$. In our extended definition of SCM — the Information Dependency Model of Definition 1 — we do not specify a precedence relation: the primitives are the information fields, and the notion of precedence is deduced from those fields. For instance, the traditional *precedence relation* on \mathbb{A} is now written as

$$\mathcal{P}a = \bigcap_{B \subset \mathbb{A}; \mathcal{J}_a \subset \mathcal{H}_B} B \quad \text{or, equivalently, } \mathcal{J}_a \subset \mathcal{H}_B \iff \mathcal{P}a \subset B. \quad (5)$$

For an SCM satisfying Equation (1), using the mapping to an IDM described in Remark 2, one can check that parental and precedence relations are related by $P(a) = \mathcal{P}a$ when $P(a)$ is the smallest set such that Equation (1) is satisfied: the relation $U_a(\omega) = \lambda_a(\omega_a, U_{P(a)}(\omega))$ implies that $\mathcal{J}_a \subset \mathcal{H}_{P(a)}$; moreover, the minimality means that $P(a)$ is the smallest subset of \mathbb{A} satisfying such constraint. So if the SCM can be represented by a DAG, $b \in \mathcal{P}a$ means that there is an arrow from b to a in this DAG.

Here is how the notion of information field allows to extend the definition of precedence to conditional precedence.

Definition 4 (Conditional Precedence) *For any subset $H \subset \mathbb{H}$ of configurations, and any subset $W \subset \mathbb{A}$, we set*

$$\mathcal{P}_{W,H}a = \bigcap_{B \subset \mathbb{A}; \mathcal{J}_a \cap H \subset \mathcal{H}_{B \cup W} \cap H} B, \quad \forall a \in \mathbb{A}.$$

Then, we call $\mathcal{P}_{W,H}$ the precedence conditioned on (W, H) binary relation on \mathbb{A} defined by $b \mathcal{P}_{W,H}a \iff b \in \mathcal{P}_{W,H}a$.

Informally, H and W are elements over which we “condition” (recall that $\mathcal{J}_a \cap H$ and $\mathcal{H}_{B \cup W} \cap H$ are trace fields over H). When $W = \emptyset$ and $H = \mathbb{H}$, we get that $\mathcal{P}_{W,H} = \mathcal{P}$.

Example 5 (Recursive Information Dependency Model) *Informally an Information Dependency Model is recursive when it corresponds to a DAG, i.e. $\mathcal{P} = \mathcal{P}_{\emptyset, \mathbb{H}}$ induces a partial order.*

Remark 6 (Solvability) *When the Information Dependency Model described by (4) has been constructed from a DAG (see Remark 2), there is no question of well-posedness. Indeed, one can simulate a sample of random variables by first generating the variables that do not have parents, and then following the graph along their children. In such situation, we can equivalently say that the IDM allows for a sequential resolution, is recursive, or admits a fixed causal ordering [26].*

However, there exist IDMs that do not have a fixed causal ordering. For such nonsequential (or equivalently, nonrecursive) IDMs, we require a weaker property than sequentiality to ensure well-posedness: solvability. We discuss in more details the question of solvability in §4.4. We need in particular to exclude cases such as self-information (that is, $a \in P(a)$) and, more generally, cases where the system of equations (1) could have several solutions (consider for instance $x = y$ and $y = x$) or no solution at all.

In [23], the authors manage to summarize the three rules of do-calculus into one rule thanks to the notions of context specific independence and labeled DAGs. Our definition allows us to reproduce their approach.

Example 7 (Context Specific Independence) *In order to model spurious edges, [23] relies on so-called labeled DAGs that can be turned into a context specific DAG by removing the arcs that are deactivated (spurious) in the context of interest. In the formalism that we propose, such context is represented by a subset H of \mathbb{H} . Indeed, if we denote by $H \subset \mathbb{H}$ the context for which an arc (a, b) is deactivated (in the language of [23]), we encode this by the two properties $a \notin \mathcal{P}_{\emptyset, H}b$ and $a \in \mathcal{P}_{\emptyset, \mathbb{H} \setminus H}b$ — themselves encoded in the structure of b 's information field \mathcal{J}_b .*

For the reader familiar with [23], it is then easy to guess how we are going to model intervention variables.

Example 8 (Intervention variables) *To introduce the possibility to intervene on a variable, we use a simple procedure. Suppose we are interested in an intervention profile $\hat{\lambda}_Z$ for a subset $Z \subset \mathbb{A}$. For this purpose, we consider a new family $\{\hat{\mathcal{J}}_z\}_{z \in Z}$ of fields $\hat{\mathcal{J}}_z \subset \mathcal{H}$, and we suppose that $\hat{\lambda}_Z$ is $\hat{\mathcal{J}}_z$ -measurable, for any $z \in Z$. Then, we enrich the model as follows: (i) we introduce a new intervention variable I (see Figure 1b), equipped with $\Omega_I = \{0, 1\}$ and $\mathbb{U}_I = \{0, 1\}$, and which only has access to its private information in Ω_I ; (ii) we straightforwardly adapt the information fields for $\mathbb{A} \setminus Z$ and the probability \mathbb{P} ; (iii) we replace the information field \mathcal{J}_z by $(\{0\} \otimes \mathcal{J}_z) \cup (\{1\} \otimes \hat{\mathcal{J}}_z)$, for $z \in Z$.*

More formally, we introduce the new model $(\tilde{\mathbb{A}}, (\tilde{\Omega}, \tilde{\mathcal{F}}), \{\tilde{\mathbb{U}}_a, \tilde{\mathbb{U}}_a\}_{a \in \tilde{\mathbb{A}}}, \{\tilde{\mathcal{J}}_a\}_{a \in \tilde{\mathbb{A}}})$, where $\tilde{\mathbb{A}} =$

$\mathbb{A} \cup \{I\}$, $\tilde{\Omega} = \Omega \times \{0, 1\}$, $\tilde{\mathcal{U}}_I = \{0, 1\}$, $\tilde{\mathcal{U}}_a = \mathcal{U}_a$ for any $a \in \mathbb{A}$, and

$$\tilde{\mathcal{J}}_a = \mathcal{J}_a \otimes \{\emptyset, \mathcal{U}_I\}, \quad \forall a \in \mathbb{A} \setminus Z, \quad (6a)$$

$$\tilde{\mathcal{J}}_Z = \hat{\mathcal{J}}_Z \otimes \mathcal{U}_I, \quad \forall z \in Z, \quad (6b)$$

$$\mathcal{J}_I = \bigotimes_{a \in \mathbb{A}} \{\emptyset, \Omega_a\} \otimes \mathcal{F}_I \otimes \bigotimes_{a \in \tilde{\mathbb{A}}} \{\emptyset, \mathcal{U}_a\}. \quad (6c)$$

We also extend the probability \mathbb{P} as a product probability $\tilde{\mathbb{P}} = \mathbb{P} \otimes \mu$ on $\tilde{\Omega}$, where μ is a full support probability on $\{0, 1\}$.

3 Topological separation, independence and do-calculus

In §3.1, we introduce the new notion of topological separation. In §3.2, we prove that topological separation implies independence, which allows us to derive a unique do-calculus rule. We use the notation $\llbracket r, s \rrbracket = \{r, r+1, \dots, s-1, s\}$ for any two integers r, s such that $r \leq s$.

3.1 Definition of topological separation

We now introduce the new notion of topological separation. We refer the reader to the companion paper [9] for additional material on the subject.

For any subsets $B \subset \mathbb{A}$ and $B_j \subset \mathbb{A}$, $j \in \llbracket 1, n \rrbracket$, we write $B_1 \sqcup \dots \sqcup B_n = B$ when we have both $B_j \cap B_k = \emptyset$ for all $j \neq k$ and $B_1 \cup \dots \cup B_n = B$. We will also say that $\{B_j\}_{j \in \llbracket 1, n \rrbracket}$ is a *splitting* of B (we do not use the vocable of partition because it is not required that the subsets B_j be nonempty).

Definition 9 (Topological Separation) *Let $H \subset \mathbb{H}$ and $B, C, W \subset \mathbb{A}$. We denote by $\overline{B}^{w,H}$ the smallest subset of \mathbb{A} that contains B and its own predecessors under $\mathcal{P}_{W,H}$ (that is, $\overline{B}^{w,H} = B \cup \mathcal{P}_{W,H}B \cup \mathcal{P}_{W,H}^2B \cup \dots$).*

We say that B and C are (conditionally) topologically separated (w.r.t.³ (W, H)), denoted by $B \perp\!\!\!\perp C \mid (W, H)$, if there exists $W_B, W_C \subset W$ such that

$$W_B \sqcup W_C = W \quad \text{and} \quad \overline{B \cup W_B}^{w,H} \cap \overline{C \cup W_C}^{w,H} = \emptyset. \quad (7)$$

Further discussion on this definition is provided in Sect. 5.1. As proved in Proposition 19 (see also [27]), $\overline{B}^{w,H}$ is the *topological closure* of B under a topology induced by the relation $\mathcal{P}_{W,H}$. As stated in the introduction, we prove in [9] that the topological separation is equivalent to the d-separation on DAGs.

Observe that the condition for topological separation is on the *existence* of a splitting of the set of variables W over which we want to condition. On a DAG, when $H = \mathbb{H}$, we have

³with respect to

topological separation of B and C with respect to W when there is a splitting (W_B, W_C) of W such that the sets of ancestors of $B \cup W_B$ and $C \cup W_C$ — defined as the union of the iterates of the set-valued mapping $B \subset \mathbb{A} \rightarrow P(B) \setminus W$ on those sets — are disjoint.

We prove in [9] that d-separation and topological separation are equivalent. We think this alternative definition of d-separation is very handy even for DAGs. Indeed, (i) the splitting (W_B, W_C) is given by an explicit formula (see [9]), (ii) once the splitting is given — which is the main difficulty — it is usually much quicker to check that the ancestors sets are disjoint than checking that all the paths between B and C are blocked by W , as illustrated in the following Example 10, illustrated by Figure 2.

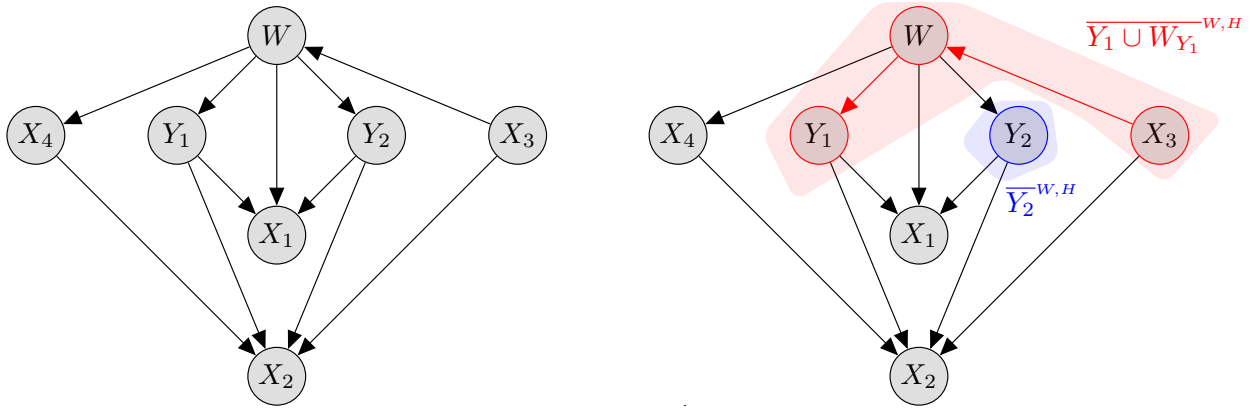
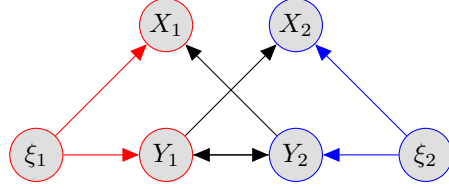
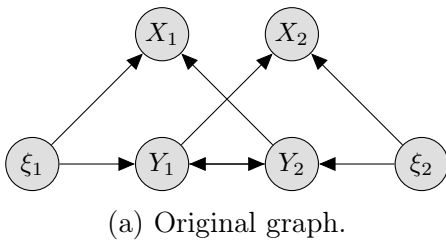


Figure 2: Topological separation w.r.t. (W, \mathbb{H}) is easy to check given the splitting $W = W_{Y_1} \sqcup W_{Y_2}$, with $W_{Y_1} = W$ and $W_{Y_2} = \emptyset$ as $\overline{Y_1 \cup W_{Y_1}}^{w,H} = \overline{Y_1 \cup W}^{w,H} = Y_1 \cup W \cup X_3$ — with, in red, the edges followed to build the closure, and the three vertices Y_1, W, X_3 where X_3 is the only new vertex in $\overline{Y_1 \cup W_{Y_1}}^{w,H} \setminus (Y_1 \cup W)$ — and $\overline{Y_2 \cup W_{Y_2}}^{w,H} = \overline{Y_2}^{w,H} = Y_2$ — with, in blue, the only vertex Y_2 because only W has an arrow pointing to Y_2 , hence has to be excluded since it is in W — do not intersect.

Example 10 (Topological separation is easy to check: recursive system) *The DAG in Figure 2 (left) illustrates why and how the notion of topological separation is practical. If one wants to check that Y_1 and Y_2 are d-separated by W , then one needs to check that every path that goes from Y_1 to Y_2 is blocked by W (by simply applying the definition), like in*

- $Y_1 \leftarrow W \rightarrow Y_2$: blocked common cause
- $Y_1 \rightarrow X_1 \leftarrow Y_2$: collider
- $Y_1 \rightarrow X_2 \leftarrow Y_2$: collider
- $Y_1 \rightarrow X_1 \leftarrow W \rightarrow Y_2$: blocked common cause
- $Y_1 \rightarrow X_2 \leftarrow X_4 \leftarrow W \rightarrow Y_2$: collider



(b) Let $W_{X_i} = Y_i$, for $i = 1, 2$. The closure of $X_1 \cup Y_1$ (resp. $X_2 \cup Y_2$), with the edges followed to build the closure, is in red (resp. blue).

Figure 3: Topological separation is easy to check: nonrecursive system.

• ...

Of course, one could simplify this long enumerating process by observing, for instance, that any path going through X_2 will be blocking, but this requires additional steps.

By contrast, the topological separation can be checked visually on the right hand side of Figure 2 by setting $W_{Y_1} = W$, $W_{Y_2} = \emptyset$ and then checking that the topological closures of $Y_1 \cup W$ (which is $Y_1 \cup W \cup X_3$) and Y_2 (which is Y_2 itself) do not intersect.

Example 11 (Topological separation is easy to check: nonrecursive system) We display in Figure 3 a nonrecursive system for which we check that X_1 and X_2 are topologically separated w.r.t. $((Y_1, Y_2), \mathbb{H})$. This is — in our opinion — simpler to check than σ -separation [12] because there are less intermediate steps.

3.2 Independence and do-calculus with information fields

In what follows, we consider random variables that take values in countable sets, and that Ω is countable as well. We pinpoint that working on continuous sets introduces technical measurability questions in the proof of Lemma 20 which are likely to be irremediable, as discussed in §5.4. We can now state our version of Pearl’s three rules of do-calculus. The statement looks like a simple sufficient condition for conditional independence thanks to the fact that we encode the intervention variables in the information fields.

Theorem 12 (Do-calculus) *Supposing that all random variables have countable codomain, and that Ω is countable as well, we have the implication*

$$Y \perp\!\!\!\perp_t Z \mid (W, H) \implies \mathbb{P}(U_Y = \cdot \mid U_W, U_{Z^{W,H}}, H) = \mathbb{P}(U_Y = \cdot \mid U_W, H) . \quad (8)$$

A more formal statement of this Theorem 12, as well as the proof and a discussion, will be provided in §5.6. We stress the conciseness of Theorem 12 – a unique rule, no “do” operator – permitted by the Information Dependency Model formulation.

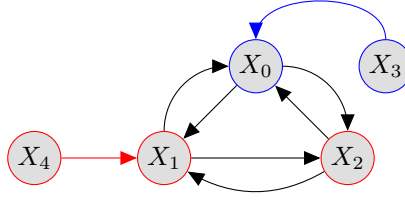


Figure 4: X_3 and X_4 are independent conditioned on (X_0, X_1, X_2) but not independent if we only condition on (X_0, X_1) . The visual proof of topological separation is obtained by considering the splitting $W_{X_4} = \{X_1, X_2\}$ and $W_{X_3} = \{X_0\}$ and observing that the topological closure of $X_3 \cup W_{X_3}$ in blue does not intersect the topological closure of $X_4 \cup W_{X_4}$ in red

Example 13 *The following, well posed (solvable as given below in Definition 15), example is inspired by the work of Witsenhausen [26] on causality. It is depicted in Figure 4 and corresponds to the following nonrecursive binary SCM (N_1, \dots, N_4 are independent, binary noise variables, \oplus is the Xor operator):*

$$\begin{aligned} X_0 &= (X_1 \cdot (\neg X_2)) \oplus (N_0 \oplus X_3) \quad \text{and} \quad X_1 = (X_2 \cdot (\neg X_0)) \oplus (N_1 \oplus X_4) \\ X_2 &= (X_0 \cdot (\neg X_1)) \oplus N_2, \quad X_3 = N_3 \quad \text{and} \quad X_4 = N_4. \end{aligned}$$

The random variables X_3 and X_4 are topologically separated by (X_0, X_1, X_2) — note that X_2 is needed — hence X_3 and X_4 are independent conditioned on (X_0, X_1, X_2) but not independent if we only condition on (X_0, X_1) .

Observe that the intuition that we could equivalently replace X_0, X_1 and X_2 by a unique variable W is misleading: with such a change, we would get a collider $X_4 \rightarrow W \leftarrow X_3$ over which we are conditioning, which would make X_4 and X_3 non-blocked with respect to W .

Let us try to apply the elegant recent result [12, Theorem 5.2] on conditional independence in the presence of cycles. We first observe that the Directed Mixed Graph (DMG) induced by the Input/output Structural Causal Model (ioSCM) associated with our example (see [12, Definitions 2.3 and 5.1]) looks like the graph of Figure 4. Second, we observe that X_0, X_1 and X_2 belong to the same strongly connected component S (see [12]), in the sense that they are all ancestors and descendants of each other. Third, let us consider the walk $X_4 \rightarrow X_1 \leftarrow X_0 \leftarrow X_3$. According to [12, Definition 4.2], the walk $X_4 \rightarrow X_1 \leftarrow X_0 \leftarrow X_3$ is $\{X_0, X_1, X_2\}$ - σ -open because

- $X_4 \rightarrow X_1 \leftarrow X_0$ satisfies the collider definition in [12, Definition 4.2, (a)], as $X_1 \in \{X_0, X_1, X_2\}$,
- $X_1 \leftarrow X_0 \leftarrow X_3$ satisfies the left chain condition because $X_0 \in \{X_0, X_1, X_2\} \cap S$, where S is the strongly connected component of X_1 .

Hence it seems that [12, Theorem 5.2] cannot be used to state that X_3 and X_4 are independent conditioned on (X_0, X_1, X_2) .

To finish, we illustrate the claim — that is, the independence of X_3 and X_4 when conditioned on X_0 , X_1 and X_2 — with a numerical exact computation, taking the N_i as binomial variables of parameter 0.1. We solve the cycle by enumerating the 8 possible combinations of values for X_0 , X_1 and X_2 and selecting the only admissible one. The results are shown in Table 1.

This example illustrates the novelty of the IDM approach.

Table 1: Numerical results for Example 13

(a) We check numerically that X_3 and X_4 are independent when conditioned on X_0 , X_1 and X_2 by computing $\mathbb{P}(X_4 = 1|X_0, X_1, X_2, X_3)$; indeed, the last two columns are identical.

(b) We check numerically that X_3 and X_4 are not independent when conditioned on X_0 and X_1 by computing $\mathbb{P}(X_4 = 1|X_0, X_1, X_3)$; indeed, the last two columns are different (see the two underlined numbers).

X_0	X_1	X_2	$X_3 = 0$	$X_3 = 1$
0	0	0	0.012	0.012
0	0	1	0.5	0.5
0	1	0	0.5	0.5
0	1	1	0.012	0.012
1	0	0	0.012	0.012
1	0	1	0.012	0.012
1	1	0	0.5	0.5
1	1	1	0.5	0.5

X_0	X_1	$X_3 = 0$	$X_3 = 1$
0	0	0.023	0.023
0	1	<u>0.1</u>	<u>0.474</u>
1	0	0.012	0.012
1	1	0.5	0.5

4 Presentation of Witsenhausen’s product model and solvability

This Sect. 4 is devoted to a presentation of the mathematical formalism and technical machinery we rely on, which we borrow from Witsenhausen’s work [26, 27]. We start in §4.1 with Witsenhausen’s product model, followed by the notions of solvability and solution map in §4.2. It is notable that our work brings together ideas from causal statistics with ideas from decentralized control theory that also attempted to provide a definition of causality a few decades ago; this is the object of §4.3. Thus equipped, we discuss cycles and the meaning of “well-posedness” in §4.4.

4.1 Witsenhausen’s product model

Because Witsenhausen introduced his model to the control community some five decades ago [26, 27], we expect that most readers will not be familiar with it. We provide tentative

correspondences between Pearl's DAG and Witsenhausen's intrinsic model in Table 2.

	<i>Pearl</i>	<i>Witsenhausen</i>
Structure	DAG	Nature and agents decision sets, with their respective fields
Parent relation	\rightarrow node	precedence relation agent
Dependence	edge SCM	agents related by the precedence relation agents information fields
Resolution	functional relation induction	policy profiles measurable w.r.t. information fields solution map
Intervention	random variable do operator	policy profile composed with solution map change of information fields
Causal ordering	fixed	existence depends on agents information fields

Table 2: Correspondences between Pearl's DAG and Witsenhausen's intrinsic model

Definition 14 (Adapted from [26, 27]) A W-model is a collection $(\mathbb{A}, \{\mathbb{U}_a, \mathcal{U}_a\}_{a \in \mathbb{A}}, (\Omega, \mathcal{F}), \{\mathcal{J}_a\}_{a \in \mathbb{A}})$, where

- \mathbb{A} is a finite set, whose elements are called agents;
- for any $a \in \mathbb{A}$, \mathbb{U}_a is a set, the set of decisions for agent a ; \mathcal{U}_a is a field over \mathbb{U}_a ;
- Ω is a set made of states of Nature; \mathcal{F} is a field over Ω ;
- for any $a \in \mathbb{A}$, \mathcal{J}_a is a subfield of the following product field

$$\mathcal{J}_a \subset \mathcal{F} \otimes \bigotimes_{b \in \mathbb{A}} \mathcal{U}_b, \quad \forall a \in \mathbb{A} \quad (9)$$

and is called the information field of the agent a .

A *countable W-model* is a W-model where all sets $\{\mathbb{U}_a\}_{a \in \mathbb{A}}$ and Ω above are countable, equipped with the complete σ -algebras.

We recall that the *configuration space* \mathbb{H} and the *configuration field* \mathcal{H} are

$$\mathbb{H} = \Omega \times \prod_{a \in \mathbb{A}} \mathbb{U}_a, \quad \mathcal{H} = \mathcal{F} \otimes \bigotimes_{a \in \mathbb{A}} \mathcal{U}_a. \quad (10)$$

A *policy* of agent $a \in \mathbb{A}$ is a mapping

$$\lambda_a : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_a, \mathcal{U}_a) \text{ such that } \lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{J}_a. \quad (11a)$$

Hence, any policy λ_a is a mapping from configurations to decisions, which satisfies the measurability property (4), that is, any policy of agent a may only depend upon the information \mathcal{J}_a available to a . We denote by Λ_a the set of all policies of agent $a \in \mathbb{A}$. A *policy profile* λ is a collection of policies, one per agent $a \in \mathbb{A}$:

$$\lambda = \{\lambda_a\}_{a \in \mathbb{A}} \in \prod_{a \in \mathbb{A}} \Lambda_a, \quad \text{where } \Lambda_a = \{\lambda_a : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_a, \mathcal{U}_a) \mid \lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{J}_a\}, \quad \forall a \in \mathbb{A}. \quad (11b)$$

4.2 Solvability and solution map

With any policy profile $\lambda = \{\lambda_a\}_{a \in \mathbb{A}} \in \prod_{a \in \mathbb{A}} \Lambda_a$ we associate the set-valued mapping

$$\mathcal{M}_\lambda : \Omega \rightrightarrows \prod_{b \in \mathbb{A}} \mathbb{U}_b, \quad \omega \mapsto \left\{ \{u_b\}_{b \in \mathbb{A}} \in \prod_{b \in \mathbb{A}} \mathbb{U}_b \mid u_a = \lambda_a(\omega, \{u_b\}_{b \in \mathbb{A}}), \quad \forall a \in \mathbb{A} \right\}. \quad (12)$$

With this definition, we slightly reformulate below how Witsenhausen introduced solvability.

Definition 15 ([26, 27]) *The solvable (measurable) property holds true for the W-model of Definition 14 — or the W-model is said to be (measurable) solvable — when, for any policy profile $\lambda = \{\lambda_a\}_{a \in \mathbb{A}} \in \prod_{a \in \mathbb{A}} \Lambda_a$, the set-valued mapping \mathcal{M}_λ in (12) is a (measurable) mapping whose domain is Ω , that is, the cardinality of $\mathcal{M}_\lambda(\omega)$ is equal to one, for any state of nature $\omega \in \Omega$. We denote by SM the solvability measurability property.*

Thus, under solvability property, for any state of nature $\omega \in \Omega$, there exists one, and only one, decision profile $\{u_b\}_{b \in \mathbb{A}} \in \prod_{b \in \mathbb{A}} \mathbb{U}_b$ which is a solution of the closed-loop equations

$$u_a = \lambda_a(\omega, \{u_b\}_{b \in \mathbb{A}}), \quad \forall a \in \mathbb{A}. \quad (13a)$$

In case of solvability, we define the solution map

$$S_\lambda : \Omega \rightarrow \mathbb{H}, \quad S_\lambda(\omega) = (\omega, M_\lambda(\omega)) \quad (13b)$$

where $M_\lambda(\omega)$ is the unique element contained in the image set $\mathcal{M}_\lambda(\omega)$ that is, for all $\{u_b\}_{b \in \mathbb{A}} \in \prod_{b \in \mathbb{A}} \mathbb{U}_b$, $M_\lambda(\omega) = \{u_b\}_{b \in \mathbb{A}} \iff \mathcal{M}_\lambda(\omega) = \{\{u_b\}_{b \in \mathbb{A}}\}$.

Thus, when the solvability property holds true, for each state of Nature, there is a single family of decisions compatible with any given policy profile. This family is the unique solution of the closed-loop equations (13a). In some cases, these equations can be solved sequentially (where the order may depend on the state of Nature, and on the given policy profile). This is the case when causality holds true.

4.3 Causality

In his articles [26, 27], Witsenhausen introduces a notion of causality that relies on suitable configuration-orderings. Here, we introduce our own notations, as they make possible a compact formulation of the causality property.

For any finite/countable set \mathbb{D} , let $|\mathbb{D}|$ denote the cardinality of \mathbb{D} . Thus, when \mathbb{A} is finite, $|\mathbb{A}|$ denotes the cardinality of the set \mathbb{A} , that is, $|\mathbb{A}|$ is the number of agents. For $k \in \llbracket 1, |\mathbb{A}| \rrbracket$, let $\Sigma_k = \{ \kappa : \llbracket 1, k \rrbracket \rightarrow \mathbb{A} \mid \kappa \text{ is an injection} \}$ denote the set of k -orderings, that is, injective mappings from $\llbracket 1, k \rrbracket$ to \mathbb{A} . The set $\Sigma_{|\mathbb{A}|}$ is the set of *total orderings* of agents in \mathbb{A} , that is, bijective mappings from $\llbracket 1, |\mathbb{A}| \rrbracket$ to \mathbb{A} (in contrast with *partial orderings* in Σ_k for $k < |\mathbb{A}|$). We define the *set of orderings* by $\Sigma = \bigcup_{k \in \llbracket 0, |\mathbb{A}| \rrbracket} \Sigma_k$, where $\Sigma_0 = \{\emptyset\}$. For any $k \in \llbracket 1, |\mathbb{A}| \rrbracket$, any ordering $\kappa \in \Sigma_k$, and any integer $\ell \leq k$, $\kappa_{\llbracket 1, \ell \rrbracket}$ is the restriction of the ordering κ to the first ℓ integers, and we introduce the mapping $\psi_k : \Sigma_{|\mathbb{A}|} \rightarrow \Sigma_k$, $\rho \mapsto \rho_{\llbracket 1, k \rrbracket}$ which yields the restriction of any total ordering of \mathbb{A} to $\llbracket 1, k \rrbracket$. For any $k \in \llbracket 1, |\mathbb{A}| \rrbracket$, and any k -ordering $\kappa \in \Sigma_k$, we define the *range* $\|\kappa\| = \{\kappa(1), \dots, \kappa(k)\} \subset \mathbb{A}$, the *cardinality* $|\kappa| = k \in \llbracket 1, |\mathbb{A}| \rrbracket$, the *last element* $\kappa^* = \kappa(k) \in \mathbb{A}$, and the *restriction* $\kappa^- = \kappa_{\llbracket 1, k-1 \rrbracket} \in \Sigma_{k-1}$.

The following definition of causality originates from [26]. In a causal W-model, there exists a configuration-ordering with the property that when an agent is called to play — as he is the last one in an ordering — what he knows cannot depend on decisions made by agents that are not his predecessors (in the range of the ordering under consideration). For this purpose, we define, for any subset $B \subset \mathbb{A}$ of agents:

$$\mathcal{U}_B = \bigotimes_{b \in B} \mathcal{U}_b \otimes \bigotimes_{a \notin B} \{\emptyset, \mathcal{U}_a\} \subset \bigotimes_{a \in \mathbb{A}} \mathcal{U}_a, \quad (14a)$$

$$\mathcal{H}_B = \mathcal{F} \otimes \mathcal{U}_B = \mathcal{F} \otimes \bigotimes_{b \in B} \mathcal{U}_b \otimes \bigotimes_{a \notin B} \{\emptyset, \mathcal{U}_a\} \subset \mathcal{H}. \quad (14b)$$

Definition 16 ([26, 27]) *A countable W-model (as in Definition 14) is causal if there exists (at least) one causal configuration-ordering $\varphi : \mathbb{H} \rightarrow \Sigma_{|\mathbb{A}|}$, that is, with the property that*

$$\mathbb{H}_\kappa^\varphi \cap H \in \mathcal{H}_{\|\kappa^-\|}, \quad \forall H \in \mathcal{J}_{\kappa^*}, \quad \forall \kappa \in \Sigma, \quad (15)$$

where the subset $\mathbb{H}_\kappa^\varphi \subset \mathbb{H}$ of configurations is defined by

$$\mathbb{H}_\emptyset^\varphi = \mathbb{H} \text{ and } \mathbb{H}_\kappa^\varphi = \{h \in \mathbb{H} \mid \psi_{|\kappa|}(\varphi(h)) = \kappa\}, \quad \forall \kappa \in \Sigma. \quad (16)$$

The set $\mathbb{H}_\kappa^\varphi$ contains all the configurations for which the agent $\kappa(1)$ is acting first, the agent $\kappa(2)$ is acting second, \dots , till the last agent $\kappa^* = \kappa(|\kappa|)$ acting at stage $|\kappa|$. Hence, otherwise said, causality means that, once we know the first $|\kappa|$ agents, the information of the agent κ^* depends at most on the decisions of the agents in the range $\|\kappa^-\|$, as represented by the subfield (see Equation (14b))

$$\mathcal{H}_{\|\kappa^-\|} = \mathcal{F} \otimes \bigotimes_{a \in \|\kappa^-\|} \mathcal{U}_a \otimes \bigotimes_{b \notin \|\kappa^-\|} \{\emptyset, \mathcal{U}_b\} \subset \mathcal{H}. \quad (17)$$

In [26], Witsenhausen proves that causal W-models are solvable measurable (SM). *The reverse is false*: in [26, Theorem 2], Witsenhausen exhibits an example of noncausal W-model that is solvable.

4.4 Cycles and well-posedness

It is notable that the framework we develop makes it possible to deal with systems with cycles. Such systems can be useful for modeling purpose. For instance, a cyclic SCM can arise as an equilibrium state of random differential equations [5]. The foundations for structural models with cycles, which are laid out in [4] show that the existence of cycles raises well-posedness questions. In particular, for such SCMs, does the system of equations (1) has a solution? is it unique? Witsenhausen introduces a hierarchy of systems that we summarize in Figure 5 (in this hierarchy a DAG corresponds to what Witsenhausen called a sequential system). For our purpose, this hierarchy could be qualified as “too strong”, because it requires the system to admit a unique solution for all policy profiles. The solvable-measurable (SM) property in Definition 15 however can be relaxed as soon as we know more about the assignment mappings under scrutiny. For instance, in [1] a property — deadlock-freeness — weaker than causality but stronger than SM is identified. Then in [2] a relaxation of deadlock-freeness is discussed: instead of imposing a constraint that need to be satisfied for every admissible policy, the authors propose to put the strain only on a policy of interest. We leave for further work a comparison between how well-posedness is handled in [4] and in the present work.

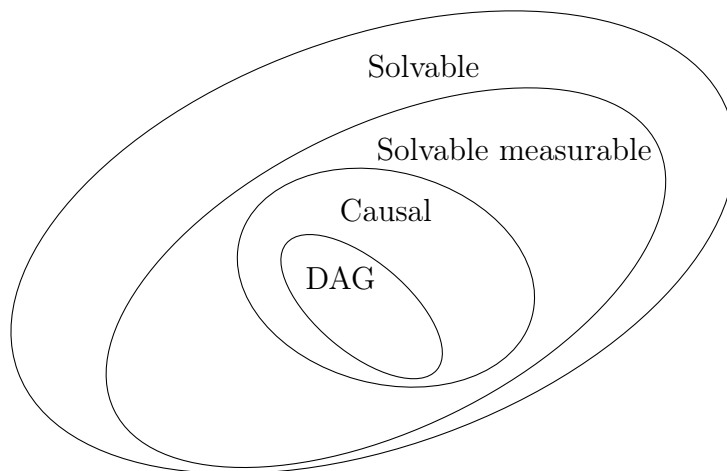


Figure 5: Hierarchy of systems

5 Formal definitions and proofs

In this Sect. 5, we provide formal definitions and proofs. In §5.1, we define conditional parentality and topological separation. We prove that topological separation implies factorization in §5.2. In §5.3, we provides tools to study conditional independence in the presence of nonrecursive systems. In §5.4, we discuss the impact of the discrete *versus* continuous settings on conditional independence. In §5.5, we show that topological separation implies

conditional independence. In §5.6, we show that the topological separation allows to define an alternative (and equivalent) do-calculus.

We provide background on binary relations. We recall that a *(binary) relation* \mathfrak{R} on \mathbb{A} is a subset $\mathfrak{R} \subset \mathbb{A}^2$ and that $a \mathfrak{R} b$ means $(a, b) \in \mathfrak{R}$. For any subset $B \subset \mathbb{A}$, the *(sub)diagonal relation* is $\Delta_B = \{(a, b) \in \mathbb{A}^2 \mid a = b \in B\}$ and the *diagonal relation* is $\Delta = \Delta_{\mathbb{A}}$. A *foreset* of a relation \mathfrak{R} is any set of the form $\mathfrak{R}b = \{a \in \mathbb{A} \mid a \mathfrak{R} b\}$, where $b \in \mathbb{A}$, or, by extension, of the form $\mathfrak{R}B = \{a \in \mathbb{A} \mid \exists b \in B, a \mathfrak{R} b\}$, where $B \subset \mathbb{A}$. The *opposite* or *complementary* \mathfrak{R}^c of a binary relation \mathfrak{R} is the relation $\mathfrak{R}^c = \mathbb{A}^2 \setminus \mathfrak{R}$, that is, defined by $a \mathfrak{R}^c b \iff \neg(a \mathfrak{R} b)$. The *converse* \mathfrak{R}^{-1} of a binary relation \mathfrak{R} is defined by $a \mathfrak{R}^{-1} b \iff b \mathfrak{R} a$ (and \mathfrak{R} is *symmetric* if $\mathfrak{R}^{-1} = \mathfrak{R}$). The *composition* $\mathfrak{R}\mathfrak{R}'$ of two binary relations $\mathfrak{R}, \mathfrak{R}'$ on \mathbb{A} is defined by $a(\mathfrak{R}\mathfrak{R}')b \iff \exists \delta \in \mathbb{A}, a \mathfrak{R} \delta$ and $\delta \mathfrak{R}' b$; then, by induction we define $\mathfrak{R}^{k+1} = \mathfrak{R}\mathfrak{R}^k$ for $k \in \mathbb{N}^*$. The *transitive closure* of a binary relation \mathfrak{R} is $\mathfrak{R}^+ = \bigcup_{k=1}^{\infty} \mathfrak{R}^k$ (and \mathfrak{R} is *transitive* if $\mathfrak{R}^+ = \mathfrak{R}$) and the *reflexive and transitive closure* is $\mathfrak{R}^* = \mathfrak{R}^+ \cup \Delta$.

5.1 Conditional parentality and topological separation

We now formally define the conditional parental relation, provide properties and then deduce a topology on the set \mathbb{A} of agents. It is this topology which is implicit in the Definition 9 of topological separation.

Definition 17 *Let $H \subset \mathbb{H}$ be a subset of configurations, and $W \subset \mathbb{A}$ be a subset of agents. We set*

$$\mathcal{P}_{W,Ha} = \bigcap_{B \subset \mathbb{A}; \mathcal{J}_a \cap H \subset \mathcal{H}_{B \cup W} \cap H} B, \quad \forall a \in \mathbb{A}, \quad (18a)$$

and we define the (conditional) parental relation $\mathcal{P}_{W,H}$ on \mathbb{A} (w.r.t. (W, H)) by

$$b \mathcal{P}_{W,H} a \iff b \in \mathcal{P}_{W,Ha}, \quad \forall (a, b) \in \mathbb{A}^2. \quad (18b)$$

We call (conditional) ancestral relation (w.r.t. (W, H)) the transitive and reflexive closure $\mathcal{P}_{W,H}^*$ of the conditional parental relation $\mathcal{P}_{W,H}$, that is,

$$\mathcal{P}_{W,H}^* = \Delta \cup \mathcal{P}_{W,H}^+ = \Delta \cup \bigcup_{k=1}^{\infty} \mathcal{P}_{W,H}^k. \quad (19)$$

Thus, when $b \mathcal{P}_{W,H} a$, it means, by (18), that the information available to agent a , on the subset $H \subset \mathbb{H}$ of configurations, involves the decisions of the agent b and, possibly of the agents in W . Witsenhausen's precedence relation \mathcal{P} in (5) is the special case $\mathcal{P}_{\emptyset, \mathbb{H}}$.

Proposition 18 *We have that*

$$\mathcal{P}_{W,H} = \Delta_{W^c} \mathcal{P}_{\emptyset, H}. \quad (20)$$

Proof. Let $a \in \mathbb{A}$ be a given agent. We introduce the two subsets of agents defined by $\Gamma_a = \{B \subset \mathbb{A} \mid \mathcal{J}_a \cap H \subset \mathcal{H}_B \cap H\}$ and $\Gamma_{a,W} = \{B \subset \mathbb{A} \mid \mathcal{J}_a \cap H \subset \mathcal{H}_{B \cup W} \cap H\}$. Then, the two subsets $\mathcal{P}_{\emptyset,Ha}$ and $\mathcal{P}_{W,Ha}$ read as $\mathcal{P}_{\emptyset,Ha} = \bigcap_{B \in \Gamma_a} B$ and $\mathcal{P}_{W,Ha} = \bigcap_{B \in \Gamma_{a,W}} B$.

As a preliminary result we prove that, for any $a \in \mathbb{A}$, we have that $\mathcal{P}_{W,Ha} \subset W^c$. Indeed, for a given agent $a \in \mathbb{A}$, we have that $J_a \subset \mathcal{H} = \mathcal{H}_{W^c \cup W}$. Thus, we obtain that $J_a \cap H \subset \mathcal{H}_{W^c \cup W} \cap H$ which gives that $W^c \in \Gamma_{a,W}$. Now, as $\mathcal{P}_{W,Ha} = \bigcap_{B \in \Gamma_{a,W}} B$, we conclude that $\mathcal{P}_{W,Ha} \subset W^c$.

Now, we establish two easy inclusions. First, for any $C \in \Gamma_{a,W}$, using the definitions of $\Gamma_{a,W}$ and Γ_a we obtain that $C \cup W \in \Gamma_a$. Thus, we have the inclusion $\{C \cup W \mid C \in \Gamma_{a,W}\} \subset \Gamma_a$ which gives

$$\bigcap_{B \in \Gamma_a} B \subset \bigcap_{C \in \Gamma_{a,W}} (C \cup W). \quad (21)$$

Second, for any $B \in \Gamma_a$, we have that $J_a \cap H \subset \mathcal{H}_B \cap H \subset \mathcal{H}_{B \cup W} \cap H = \mathcal{H}_{(B \cap W^c) \cup W} \cap H$ which gives that $B \cap W^c \in \Gamma_{a,W}$. Thus we have that $\{B \cap W^c \mid B \in \Gamma_a\} \subset \Gamma_{a,W}$, which gives

$$\bigcap_{C \in \Gamma_{a,W}} C \subset \bigcap_{B \in \Gamma_a} (B \cap W^c). \quad (22)$$

Now, we successively have

$$\begin{aligned} W^c \cap (\mathcal{P}_{\emptyset,Ha}) &= W^c \cap \left(\bigcap_{B \in \Gamma_a} B \right) && \text{(as } \mathcal{P}_{\emptyset,Ha} = \bigcap_{B \in \Gamma_a} B \text{)} \\ &\subset W^c \cap \left(\bigcap_{C \in \Gamma_{a,W}} (C \cup W) \right) && \text{(by (21))} \\ &= W^c \cap \left(\left(\bigcap_{C \in \Gamma_{a,W}} C \right) \cup W \right) \\ &= \left(W^c \cap \left(\bigcap_{C \in \Gamma_{a,W}} C \right) \right) \cup (W^c \cap W) \\ &= W^c \cap (\mathcal{P}_{W,Ha}) && \text{(as } \mathcal{P}_{W,Ha} = \bigcap_{C \in \Gamma_{a,W}} C \text{)} \\ &= \mathcal{P}_{W,Ha} && \text{(by preliminary result } \mathcal{P}_{W,Ha} \subset W^c \text{)} \\ &\subset \bigcap_{B \in \Gamma_a} (B \cap W^c) && \text{(by (22) as } \mathcal{P}_{W,Ha} = \bigcap_{C \in \Gamma_{a,W}} C \text{)} \\ &= W^c \cap (\mathcal{P}_{\emptyset,Ha}). \end{aligned}$$

Therefore, we have obtained that $W^c \cap (\mathcal{P}_{\emptyset,Ha}) = \mathcal{P}_{W,Ha}$, that is, $\Delta_{W^c} \mathcal{P}_{\emptyset,H} = \mathcal{P}_{W,H}$. \square

In [27], Witsenhausen introduced a topology on the set \mathbb{A} of agents related to the precedence relation \mathcal{P} in (5). Here, we extend his approach to the conditional parental relation $\mathcal{P}_{W,H}$ on \mathbb{A} .

Proposition 19 *There exists a topology on the set \mathbb{A} of agents such that the topological closure $\overline{B}^{W,H}$ of a subset $B \subset \mathbb{A}$ is the $\mathcal{P}_{W,H}^*$ -foreset*

$$\overline{B}^{W,H} = \mathcal{P}_{W,H}^* B. \quad (23)$$

In this topology, the subset W is open or, equivalently, the subset W^c is closed.

We refer the reader to the companion paper [9] for additional material on the aforementioned topology. For the sake of completeness, we give a proof. We mention that $\overline{B}^{w,h}$ is the smallest subset that contains B such that $\mathcal{P}_{W,H}\overline{B}^{w,h} \subset \overline{B}^{w,h}$.

Proof. We define $\mathcal{F}^{w,h} = \{C \subset \mathbb{A} \mid \mathcal{P}_{W,H}^*C \subset C\}$ and we show that the elements of $\mathcal{F}^{w,h}$ form the closed sets of a topology on the set \mathbb{A} . In fact, we are going to prove the stronger property that the set $\mathcal{F}^{w,h}$ is an Alexandrov topology: it contains both \emptyset, \mathbb{A} and is stable under the union and intersection operations (not necessarily finite). Indeed, both $\emptyset, \mathbb{A} \in \mathcal{F}^{w,h}$ as $\mathcal{P}_{W,H}^*\emptyset = \emptyset$ and $\mathcal{P}_{W,H}^*\mathbb{A} \subset \mathbb{A}$. Moreover, let $\{B_s\}_{s \in \mathbb{S}}$ be family of elements $B_s \in \mathcal{F}^{w,h}$, that is, $\mathcal{P}_{W,H}^*B_s \subset B_s$, for all $s \in \mathbb{S}$. We have $\mathcal{P}_{W,H}^*(\bigcup_{s \in \mathbb{S}} B_s) = \bigcup_{s \in \mathbb{S}} \mathcal{P}_{W,H}^*B_s \subset \bigcup_{s \in \mathbb{S}} B_s$, hence stability by union. We have $\mathcal{P}_{W,H}^*(\bigcap_{s \in \mathbb{S}} B_s) \subset \bigcap_{s \in \mathbb{S}} \mathcal{P}_{W,H}^*B_s \subset \bigcap_{s \in \mathbb{S}} B_s$, hence stability by intersection. Thus, we have shown that $\mathcal{F}^{w,h}$ is an Alexandrov topology.

We prove Equation (23). By definition of the $\mathcal{F}^{w,h}$ topology, a subset $B \subset \mathbb{A}$ is closed iff $\mathcal{P}_{W,H}^*B \subset B$. This is also equivalent to $\mathcal{P}_{W,H}^*B = B$ because $B \subset \mathcal{P}_{W,H}^*B$ since the relation $\mathcal{P}_{W,H}^* = (\mathcal{P}_{W,H})^+ \cup \Delta$ is reflexive. Now, we consider a subset $B \subset \mathbb{A}$ and we characterize its topological closure $\overline{B}^{w,h}$, the smallest closed subset that contains B . On the one hand, we have that $B \subset \mathcal{P}_{W,H}^*B$ because the relation $\mathcal{P}_{W,H}^* = (\mathcal{P}_{W,H})^+ \cup \Delta$ is reflexive. On the other hand, the set $\mathcal{P}_{W,H}^*B$ is closed since $\mathcal{P}_{W,H}^*(\mathcal{P}_{W,H}^*B) = (\mathcal{P}_{W,H}^*)^2B \subset \mathcal{P}_{W,H}^*B$, because the relation $\mathcal{P}_{W,H}^* = (\mathcal{P}_{W,H})^+ \cup \Delta$ is transitive. By definition of the topological closure $\overline{B}^{w,h}$ we deduce that $\overline{B}^{w,h} \subset \mathcal{P}_{W,H}^*B$. Now, we prove the reverse inclusion. Let $C \subset \mathbb{A}$ be a closed subset such that $B \subset C$, we necessarily have that $\mathcal{P}_{W,H}^*B \subset \mathcal{P}_{W,H}^*C \subset C$ and thus $\mathcal{P}_{W,H}^*B \subset C$. Now, considering the special case where $C = \overline{B}^{w,h}$ which is a closed subset containing B we obtain that $\mathcal{P}_{W,H}^*B \subset \overline{B}^{w,h}$. We conclude that $\mathcal{P}_{W,H}^*B = \overline{B}^{w,h}$.

The subset W is open because its complementary set W^c satisfies $\mathcal{P}_{W,H}^*W^c = (\mathcal{P}_{W,H})^+W^c \cup W^c \subset W^c$, as $\mathcal{P}_{W,H}\mathbb{A} \subset W^c$ because $\mathcal{P}_{W,H} = \Delta_{W^c}\mathcal{P}_{\emptyset,H}$ by (20) and by definition of the subdiagonal relation Δ_{W^c} .

This ends the proof. □

5.2 Topological separation implies factorization

From now on, as we deal for the first time with probability, we consider a countable W-model (as in Definition 14), that is, a W-model where all sets \mathbb{A} , $\{\mathbb{U}_a\}_{a \in \mathbb{A}}$ and Ω are countable, equipped with the complete σ -algebras. Moreover, we suppose that the set Ω of states of Nature, and its field \mathcal{F} have the following product form:

$$\Omega = \prod_{a \in \mathbb{A}} \Omega_a, \quad \mathcal{F} = \bigotimes_{a \in \mathbb{A}} \mathcal{F}_a. \quad (24)$$

For any nonempty subset $B \subset \mathbb{A}$ of agents, we denote

$$\Omega_B = \prod_{b \in B} \Omega_b, \quad \mathcal{F}_B = \bigotimes_{b \in B} \mathcal{F}_b \quad (25a)$$

$$h_B = \{h_b\}_{b \in B} \in \prod_{b \in B} \mathbb{U}_b, \quad \forall h \in \mathbb{H} = \Omega \times \prod_{a \in \mathbb{A}} \mathbb{U}_a, \quad (25b)$$

$$\pi_B : \mathbb{H} \rightarrow \prod_{b \in B} \mathbb{U}_b, \quad h \mapsto h_B, \quad (25c)$$

$$\lambda_B = \{\lambda_b\}_{b \in B} \in \prod_{b \in B} \Lambda_b, \quad \forall \lambda \in \Lambda. \quad (25d)$$

We are now going to show how conditional topological separation induces a factorization of the solution map (see Definition 15).

Lemma 20 *We consider a solvable countable W -model, where the set Ω of states of Nature has the product form (24), where each information field \mathcal{J}_a in (9) is such that*

$$\mathcal{J}_a \subset \mathcal{F}_a \otimes \bigotimes_{b \neq a} \{\emptyset, \Omega_b\} \times \bigotimes_{c \in \mathbb{A}} \mathcal{U}_c, \quad \forall a \in \mathbb{A}. \quad (26)$$

We also consider a policy profile $\lambda = \{\lambda_a\}_{a \in \mathbb{A}} \in \prod_{a \in \mathbb{A}} \Lambda_a$, a subset $H \subset \mathbb{H}$ of configurations, and Y, W and Z three subsets of \mathbb{A} , two by two disjoint and such that (see Definition 9 for the notation $\perp\!\!\!\perp_t$)

$$Y \perp\!\!\!\perp_t Z \mid (W, H). \quad (27)$$

Then, there exist five subsets $Y', Z', W_Y, W_Z, E \subset \mathbb{A}$ such that

$$\mathbb{A} = \tilde{Y} \sqcup \tilde{Z} \sqcup E \quad \text{where} \quad \tilde{Y} = Y \sqcup Y' \sqcup W_Y, \quad \tilde{Z} = Z \sqcup Z' \sqcup W_Z, \quad W = W_Y \sqcup W_Z, \quad (28a)$$

and there exist three measurable mappings (reduced solution maps)

$$\tilde{M}_{\lambda_{\tilde{Y}}} : \Omega_{\tilde{Y}} \times \mathbb{U}_{W_Z} \rightarrow \mathbb{U}_{\tilde{Y}}, \quad \tilde{M}_{\lambda_{\tilde{Z}}} : \Omega_{\tilde{Z}} \times \mathbb{U}_{W_Y} \rightarrow \mathbb{U}_{\tilde{Z}}, \quad \tilde{M}_{\lambda_E} : \Omega_E \times \mathbb{U}_{\tilde{Y} \cup \tilde{Z}} \rightarrow \mathbb{U}_E \quad (28b)$$

such that the solution map $S_\lambda(\omega) = (\omega, M_\lambda(\omega))$ in (13b) splits in three factors as follows: $\forall \omega \in S_\lambda^{-1}(H)$, we have that

$$M_\lambda(\omega) = \left(\tilde{M}_{\lambda_{\tilde{Y}}}(\omega_{\tilde{Y}}, \lambda_{W_Z}(S_\lambda(\omega))), \tilde{M}_{\lambda_{\tilde{Z}}}(\omega_{\tilde{Z}}, \lambda_{W_Y}(S_\lambda(\omega))), \tilde{M}_{\lambda_E}(\omega_E, \lambda_{\tilde{Y} \cup \tilde{Z}}(S_\lambda(\omega))) \right), \quad (28c)$$

or, equivalently, with the notation (25c),

$$M_\lambda(\omega) = \left(\tilde{M}_{\lambda_{\tilde{Y}}}(\omega_{\tilde{Y}}, \pi_{W_Z}(S_\lambda(\omega))), \tilde{M}_{\lambda_{\tilde{Z}}}(\omega_{\tilde{Z}}, \pi_{W_Y}(S_\lambda(\omega))), \tilde{M}_{\lambda_E}(\omega_E, \pi_{\tilde{Y} \cup \tilde{Z}}(S_\lambda(\omega))) \right). \quad (28d)$$

More precisely, again with the notation (25c), Equation (28c) has to be understood as $\pi_{\tilde{Y}}(M_\lambda(\omega)) = \tilde{M}_{\lambda_{\tilde{Y}}}(\omega_{\tilde{Y}}, \lambda_{W_Z}(S_\lambda(\omega)))$, $\pi_{\tilde{Z}}(M_\lambda(\omega)) = \tilde{M}_{\lambda_{\tilde{Z}}}(\omega_{\tilde{Z}}, \lambda_{W_Y}(S_\lambda(\omega)))$, and $\pi_E(M_\lambda(\omega)) \equiv \tilde{M}_{\lambda_E}(\omega_E, \lambda_{\tilde{Y} \cup \tilde{Z}}(S_\lambda(\omega)))$.

Before providing the proof we introduce the following preliminary result, which is an application of a result by Doob (see [10, Theorem 18 in Chapter 2] and [11]) in a countable setting.

Lemma 21 *We consider a countable W -model. Let $A \subset \mathbb{A}$ and $B \subset \mathbb{A}$. Let π be the projection mapping from \mathbb{H} to $\Omega_A \times \mathbb{U}_B$, and λ_A a policy profile for the elements of A . Let $H \subset \mathbb{H}$ be such that $\sigma(\lambda_A) \cap H \subseteq \sigma(\pi) \cap H$. Then, there exist a mapping $\hat{\lambda}_A : \Omega_A \times \mathbb{U}_B \rightarrow \mathbb{U}_A$ such that $\lambda_A^H = \hat{\lambda} \circ \pi^H$, where λ_A^H and π^H are the restrictions of λ_A and π to H .*

Proof. This is a trivial application of Doob Lemma on a discrete set. First we observe that the hypothesis implies that $\sigma(\lambda_A^H) \subset \sigma(\pi^H)$. As the set \mathbb{H} is countable, so is the set H and we can apply Doob's Theorem [10, Theorem 18 in Chapter 2], which implies that there exists a mapping $\hat{\lambda} : \text{im}(\pi^H) \subset \Omega_A \times \mathbb{U}_B \rightarrow \mathbb{U}_A$ such that $\lambda_A^H = \hat{\lambda} \circ \pi^H$ (where $\text{im}(\pi^H)$ is the the image of π^H). We can then extend the domain of $\hat{\lambda}$, so that $\hat{\lambda} : \Omega_A \times \mathbb{U}_B \rightarrow \mathbb{U}_A$ is such that $\lambda_A^H = \hat{\lambda} \circ \pi^H$ which is what we wanted to show. \square

Proof. The proof is in five steps.

- First, we identify five subsets $Y', Z', W_Y, W_Z, E \subset \mathbb{A}$ such that (28a) holds true.

By assumption, we have that $Y \cap Z = Y \cap W = Z \cap W = \emptyset$ and $Y \perp\!\!\!\perp_t Z \mid (W, H)$. As a consequence, by Definition 9, there exists $W_Y, W_Z \subset W$ such that $W_Y \sqcup W_Z = W$ and $\overline{Y \cup W_Y}^{w, H} \cap \overline{Z \cup W_Z}^{w, H} = \emptyset$, that is,

$$W_Y \cap W_Z = \emptyset, \quad W_Y \cup W_Z = W, \quad \mathcal{P}_{W, H}^*(Y \cup W_Y) \cap \mathcal{P}_{W, H}^*(Z \cup W_Z) = \emptyset. \quad (29a)$$

We set

$$\tilde{Y} = \mathcal{P}_{W, H}^*(Y \cup W_Y) \quad \text{and} \quad \tilde{Z} = \mathcal{P}_{W, H}^*(Z \cup W_Z). \quad (29b)$$

By definition of the ancestral relation $\mathcal{P}_{W, H}^* = \mathcal{P}_{W, H}^+ \cup \Delta$ in (19), we have that $Y \cup W_Y \subset \tilde{Y}$ and $Z \cup W_Z \subset \tilde{Z}$, where we can write $Y \cup W_Y = Y \sqcup W_Y$ and $Z \cup W_Z = Z \sqcup W_Z$ because $Y \cap W_Y \subset Y \cap W = \emptyset$ and $Z \cap W_Z \subset Z \cap W = \emptyset$ by assumption. Then, we set

$$Y' = \tilde{Y} \setminus (Y \sqcup W_Y), \quad Z' = \tilde{Z} \setminus (Z \sqcup W_Z).$$

- Second, we show that

$$\mathcal{J}_{\tilde{Y}} \cap H \subset \mathcal{H}_{\tilde{Y} \cup W_Z} \cap H \quad \text{and} \quad \mathcal{J}_{\tilde{Z}} \cap H \subset \mathcal{H}_{\tilde{Z} \cup W_Y} \cap H. \quad (29c)$$

Indeed, we have that

$$\begin{aligned} \mathcal{P}_{W, H} \tilde{Y} &= \mathcal{P}_{W, H} \mathcal{P}_{W, H}^*(Y \cup W_Y) && \text{(as } \tilde{Y} = \mathcal{P}_{W, H}^*(Y \cup W_Y) \text{ by definition (29b))} \\ &\subset \mathcal{P}_{W, H}^*(Y \cup W_Y) = \tilde{Y}. && \text{(by definition (29b))} \end{aligned}$$

Therefore, using the fact that $\mathcal{P}_{W, H} A \subset B \iff \mathcal{J}_A \cap H \subset \mathcal{H}_{B \cup W} \cap H$ by definition of $\mathcal{P}_{W, H}$, we get that

$$\mathcal{J}_{\tilde{Y}} \cap H \subset \mathcal{H}_{\tilde{Y} \cup W} \cap H,$$

which, combined with the equality $\tilde{Y} \cup W = \mathcal{P}_{W,H}^*(Y \cup W_Y) \cup W_Y \cup W_Z = \mathcal{P}_{W,H}^*(Y \cup W_Y) \cup W_Z = \tilde{Y} \cup W_Z$ gives

$$\mathcal{J}_{\tilde{Y}} \cap H \subset \mathcal{H}_{\tilde{Y} \cup W_Z} \cap H .$$

In the same way, we obtain that $\mathcal{J}_{\tilde{Z}} \cap H \subset \mathcal{H}_{\tilde{Z} \cup W_Y} \cap H$.

• Third, we prepare the existence of a factorization as in (28c). Using Equations (29a) and (29b) we have that $\tilde{Y} \cap \tilde{Z} = \emptyset$. We define $E = \mathbb{A} \setminus (\tilde{Y} \sqcup \tilde{Z})$ to obtain the decomposition $\mathbb{A} = \tilde{Y} \sqcup \tilde{Z} \sqcup E$. The solution map (13b) splits in three factors (where the projection π has been introduced in (25c))

$$M_\lambda(\omega) = \left(\pi_{\tilde{Y}}(S_\lambda(\omega)), \pi_{\tilde{Z}}(S_\lambda(\omega)), \pi_E(S_\lambda(\omega)) \right) .$$

Let us examine the term $\pi_{\tilde{Y}}(S_\lambda(\omega))$, as the other two terms can be treated in the same way. On the one hand, by (26), we have that $\mathcal{J}_{\tilde{Y}} \subset \mathcal{F}_{\tilde{Y}} \otimes \bigotimes_{b \notin \tilde{Y}} \{\emptyset, \Omega_b\} \otimes \bigotimes_{c \in \mathbb{A}} \mathcal{U}_c$. On the other hand, by (29c),

we have that $\mathcal{J}_{\tilde{Y}} \cap H \subset \mathcal{H}_{\tilde{Y} \cup W_Z} \cap H = \mathcal{F} \otimes \bigotimes_{b \in \tilde{Y} \cup W_Z} \mathcal{U}_b \otimes \bigotimes_{c \notin \tilde{Y} \cup W_Z} \{\emptyset, \mathbb{U}_c\} \cap H$. Therefore, we deduce that

$$\begin{aligned} \mathcal{J}_{\tilde{Y}} \cap H &\subset \left(\left(\mathcal{F}_{\tilde{Y}} \otimes \bigotimes_{b \notin \tilde{Y}} \{\emptyset, \Omega_b\} \otimes \bigotimes_{c \in \mathbb{A}} \mathcal{U}_c \right) \cap \left(\mathcal{F} \otimes \bigotimes_{b \in \tilde{Y} \cup W_Z} \mathcal{U}_b \otimes \bigotimes_{c \notin \tilde{Y} \cup W_Z} \{\emptyset, \mathbb{U}_c\} \right) \right) \cap H \\ &= \left(\mathcal{F}_{\tilde{Y}} \otimes \bigotimes_{b \notin \tilde{Y}} \{\emptyset, \Omega_b\} \otimes \bigotimes_{b \in \tilde{Y} \cup W_Z} \mathcal{U}_b \otimes \bigotimes_{c \notin \tilde{Y} \cup W_Z} \{\emptyset, \mathbb{U}_c\} \right) \cap H . \end{aligned}$$

By Lemma 21, there exists a “reduced” mapping $\bar{\lambda}_{\tilde{Y}} : \Omega_{\tilde{Y}} \times \mathbb{U}_{\tilde{Y}} \times \mathbb{U}_{W_Z} \rightarrow \mathbb{U}_{\tilde{Y}}$ such that

$$\begin{aligned} \lambda_{\tilde{Y}}(\omega_{\tilde{Y}}, \omega_{\mathbb{A} \setminus \tilde{Y}}, u_{\tilde{Y}}, u_{W_Z}, u_{\mathbb{A} \setminus (\tilde{Y} \cup W_Z)}) &= \bar{\lambda}_{\tilde{Y}}(\omega_{\tilde{Y}}, u_{\tilde{Y}}, u_{W_Z}) , \\ \forall (\omega_{\tilde{Y}}, \omega_{\mathbb{A} \setminus \tilde{Y}}, u_{\tilde{Y}}, u_{W_Z}, u_{\mathbb{A} \setminus (\tilde{Y} \cup W_Z)}) &\in H . \end{aligned}$$

In the same way, there exists a “reduced” mapping $\bar{\lambda}_{\tilde{Z}} : \Omega_{\tilde{Z}} \times \mathbb{U}_{\tilde{Z}} \times \mathbb{U}_{W_Y} \rightarrow \mathbb{U}_{\tilde{Z}}$ such that

$$\begin{aligned} \lambda_{\tilde{Z}}(\omega_{\tilde{Z}}, \omega_{\mathbb{A} \setminus \tilde{Z}}, u_{\tilde{Z}}, u_{W_Y}, u_{\mathbb{A} \setminus (\tilde{Z} \cup W_Y)}) &= \bar{\lambda}_{\tilde{Z}}(\omega_{\tilde{Z}}, u_{\tilde{Z}}, u_{W_Y}) , \\ \forall (\omega_{\tilde{Z}}, \omega_{\mathbb{A} \setminus \tilde{Z}}, u_{\tilde{Z}}, u_{W_Y}, u_{\mathbb{A} \setminus (\tilde{Z} \cup W_Y)}) &\in H , \end{aligned}$$

and a mapping $\bar{\lambda}_E : \Omega_E \times \mathbb{U}_E \times \mathbb{U}_{E^c} \rightarrow \mathbb{U}_E$ such that

$$\lambda_E(\omega_E, \omega_{E^c}, u_E, u_{E^c}) = \bar{\lambda}_E(\omega_E, u_E, u_{E^c}) , \quad \forall (\omega_E, \omega_{E^c}, u_E, u_{E^c}) \in H .$$

By Definition 15 of $S_\lambda(\omega)$, we can regroup — for $\omega \in S_\lambda^{-1}(H)$ — the closed-loop equations (13a) in three parts as

$$\begin{aligned} \pi_{\tilde{Y}}(S_\lambda(\omega)) &= \bar{\lambda}_{\tilde{Y}}(\omega_{\tilde{Y}}, \pi_{\tilde{Y}}(S_\lambda(\omega)), \pi_{W_Z}(S_\lambda(\omega))) , \\ \pi_{\tilde{Z}}(S_\lambda(\omega)) &= \bar{\lambda}_{\tilde{Z}}(\omega_{\tilde{Z}}, \pi_{\tilde{Z}}(S_\lambda(\omega)), \pi_{W_Y}(S_\lambda(\omega))) , \\ \pi_E(S_\lambda(\omega)) &= \bar{\lambda}_E(\omega_E, \pi_E(S_\lambda(\omega)), \pi_{\tilde{Y}}(S_\lambda(\omega)), \pi_{\tilde{Z}}(S_\lambda(\omega))) , \end{aligned}$$

so that — for $\omega \in S_\lambda^{-1}(H)$ — the reduced closed-loop equations

$$u_{\tilde{Y}} = \bar{\lambda}_{\tilde{Y}}(\omega_{\tilde{Y}}, u_{\tilde{Y}}, u_{W_Z}) , \quad (30a)$$

$$u_{\tilde{Z}} = \bar{\lambda}_{\tilde{Z}}(\omega_{\tilde{Z}}, u_{\tilde{Z}}, u_{W_Y}) , \quad (30b)$$

$$u_E = \bar{\lambda}_E(\omega_E, u_E, u_{\tilde{Y}}, u_{\tilde{Z}}) , \quad (30c)$$

have (at least) the solution $(u_{\tilde{Y}}, u_{\tilde{Z}}, u_E) = \left(\pi_{\tilde{Y}}(S_\lambda(\omega)), \pi_{\tilde{Z}}(S_\lambda(\omega)), \pi_E(S_\lambda(\omega)) \right)$ when $u_W = \pi_W(S_\lambda(\omega))$.

• Fourth, we show the existence of three mappings as in (28b).

Let $\omega \in S_\lambda^{-1}(H)$. We denote by $\mathbb{U}_{\tilde{Y}}(\omega)$ the set of elements $u_{\tilde{Y}} \in \mathbb{U}_{\tilde{Y}}$ such that there exists at least one $(\omega'_{\tilde{Z}}, \omega'_E, u_{\tilde{Z}}, u_E) \in \Omega_{\tilde{Z}} \times \Omega_E \times \mathbb{U}_{\tilde{Z}} \times \mathbb{U}_E$ that satisfies $(u_{\tilde{Y}}, u_{\tilde{Z}}, u_E) = \lambda((\omega_{\tilde{Y}}, \omega'_{\tilde{Z}}, \omega'_E), (u_{\tilde{Y}}, u_{\tilde{Z}}, u_E))$ and $\pi_{W_Z}(S_\lambda(\omega)) = \pi_{W_Z}(S_\lambda(\omega_{\tilde{Y}}, \omega'_{\tilde{Z}}, \omega'_E))$. We are going to show that $\mathbb{U}_{\tilde{Y}}(\omega)$ is a singleton. For this purpose, we consider $(u_{\tilde{Y}}, \omega_{\tilde{Z}}, \omega_E, u_{\tilde{Z}}, u_E)$ and $(u'_{\tilde{Y}}, \omega'_{\tilde{Z}}, \omega'_E, u'_{\tilde{Z}}, u'_E)$ satisfying the two conditions that define the set $\mathbb{U}_{\tilde{Y}}(\omega)$.

As we have, on the one hand, that

$$(u_{\tilde{Y}}, u_{\tilde{Z}}, u_E) = \lambda((\omega_{\tilde{Y}}, \omega_{\tilde{Z}}, \omega_E), (u_{\tilde{Y}}, u_{\tilde{Z}}, u_E)) \quad (31a)$$

and, on the other hand, that

$$(u'_{\tilde{Y}}, u'_{\tilde{Z}}, u'_E) = \lambda((\omega_{\tilde{Y}}, \omega'_{\tilde{Z}}, \omega'_E), (u'_{\tilde{Y}}, u'_{\tilde{Z}}, u'_E)) , \quad (31b)$$

we deduce that, by using (30) and the condition on u_{W_Z} in the definition of $\mathbb{U}_{\tilde{Y}}(\omega)$:

$$(u_{\tilde{Y}}, u'_{\tilde{Z}}, \hat{u}_E) = \lambda((\omega_{\tilde{Y}}, \omega'_{\tilde{Z}}, \omega'_E), (u_{\tilde{Y}}, u'_{\tilde{Z}}, \hat{u}_E)) , \quad (31c)$$

where $\hat{u}_E = \bar{\lambda}_E(\omega'_E, \hat{u}_E, \lambda_{\tilde{Y} \cup \tilde{Z}}(S_\lambda(\omega)))$. The Equations (31b) and (31c) imply, by the solvability assumption (see Definition 15), that $u_{\tilde{Y}} = u'_{\tilde{Y}}$. As a consequence, we have proven that the set $\mathbb{U}_{\tilde{Y}}(\omega)$ is a singleton.

Thus, we have defined, for any $u_{W_{\tilde{Z}}} = \pi_{W_{\tilde{Z}}}(S_\lambda(\omega))$ a unique element $u_{\tilde{Y}} = \widetilde{M}_{\lambda_{\tilde{Y}}}(\omega_{\tilde{Y}}, u_{W_{\tilde{Z}}})$. We do the same for \tilde{Z} and for E . Thus, we have defined reduced solution maps as follows

$$\widetilde{M}_{\lambda_{\tilde{Y}}} : \Omega_{\tilde{Y}} \times \mathbb{U}_{W_Z} \rightarrow \mathbb{U}_{\tilde{Y}} , \quad (32a)$$

$$\widetilde{M}_{\lambda_{\tilde{Z}}} : \Omega_{\tilde{Z}} \times \mathbb{U}_{W_Y} \rightarrow \mathbb{U}_{\tilde{Z}} , \quad (32b)$$

$$\widetilde{M}_{\lambda_E} : \Omega_E \times \mathbb{U}_{\tilde{Y} \cup \tilde{Z}} \rightarrow \mathbb{U}_E . \quad (32c)$$

As we considered that all sets \mathbb{A} , $\{\mathbb{U}_a\}_{a \in \mathbb{A}}$ and Ω are countable, the above mappings are measurable.

This ends the proof. □

5.3 Conditional independence in the presence of cycles

This subsection provides tools to study conditional independence in the presence of nonrecursive systems. We also discuss an instance where such independence is not captured by Pearl's d-separation criterion [16].

5.3.1 Key technical lemma for dealing with cycles

We state and prove a lemma that will serve as a main argument for the proof of the coming Theorem 24. As far as we know, this result is novel. It cannot be deduced from Pearl's rules.

Lemma 22 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\Xi_1, \Xi_2, \Upsilon_1, \Upsilon_2, \Theta_1, \Theta_2$ be six measurable spaces and*

$$\Psi_1 : \Xi_1 \times \Upsilon_2 \rightarrow \Theta_1, \quad \Psi_2 : \Xi_2 \times \Upsilon_1 \rightarrow \Theta_2, \quad \Phi_1 : \Xi_1 \times \Upsilon_2 \rightarrow \Upsilon_1, \quad \Phi_2 : \Xi_2 \times \Upsilon_1 \rightarrow \Upsilon_2 \quad (33)$$

be four measurable mappings. Let

$$\xi_1 : \Omega \rightarrow \Xi_1, \quad \xi_2 : \Omega \rightarrow \Xi_2, \quad \theta_1 : \Omega \rightarrow \Theta_1, \quad \theta_2 : \Omega \rightarrow \Theta_2, \quad v_1 : \Omega \rightarrow \Upsilon_1, \quad v_2 : \Omega \rightarrow \Upsilon_2 \quad (34)$$

be six random variables satisfying

$$\theta_1 = \Psi_1(\xi_1, v_2), \quad (35a)$$

$$\theta_2 = \Psi_2(\xi_2, v_1), \quad (35b)$$

$$v_1 = \Phi_1(\xi_1, v_2), \quad (35c)$$

$$v_2 = \Phi_2(\xi_2, v_1). \quad (35d)$$

Suppose that the couple (v_1, v_2) of random variables takes values in a countable product subset $\Upsilon'_1 \times \Upsilon'_2 \subset \Upsilon_1 \times \Upsilon_2$, and that the system of equations

$$w_1 = \Phi_1(x_1, w_2), \quad (36a)$$

$$w_2 = \Phi_2(x_2, w_1), \quad (36b)$$

has a unique solution (w_1, w_2) in $\Upsilon'_1 \times \Upsilon'_2$, for any $(x_1, x_2) \in \Xi_1 \times \Xi_2$.

Then, if the random variables ξ_1 and ξ_2 are independent, the random variables θ_1 and θ_2 are independent when conditioned on (v_1, v_2) .

Proof. By assumption, there exists a unique solution $(w_1, w_2) \in \Upsilon'_1 \times \Upsilon'_2$ to the implicit system (36) of equations. Thus, there exist mappings

$$\tilde{\Phi}_1 : \Xi_1 \times \Xi_2 \rightarrow \Upsilon'_1, \quad \tilde{\Phi}_2 : \Xi_1 \times \Xi_2 \rightarrow \Upsilon'_2, \quad (37a)$$

such that, for any (w_1, w_2) in $\Upsilon'_1 \times \Upsilon'_2$ and $(x_1, x_2) \in \Xi_1 \times \Xi_2$, we have

$$\left(w_1 = \Phi_1(x_1, w_2), \quad w_2 = \Phi_2(x_2, w_1) \right) \iff \left(w_1 = \tilde{\Phi}_1(x_1, x_2), \quad w_2 = \tilde{\Phi}_2(x_1, x_2) \right). \quad (37b)$$

We suppose that the random variables ξ_1 and ξ_2 are independent. We are going to show that the random variables θ_1 and θ_2 are independent when conditioned on (v_1, v_2) .

• First, we establish that, for any couple $(w_1, w_2) \in \Upsilon'_1 \times \Upsilon'_2$:

$$\left\{ \Phi_1(\xi_1, w_2) = w_1, \quad \Phi_2(\xi_2, w_1) = w_2 \right\} = \left\{ v_1 = w_1, \quad v_2 = w_2 \right\}. \quad (38)$$

Indeed, on the one hand, we have

$$\begin{aligned}
& \left\{ \Phi_1(\xi_1, w_2) = w_1, \Phi_2(\xi_2, w_1) = w_2 \right\} \\
&= \left\{ w_1 = \tilde{\Phi}_1(\xi_1, \xi_2), w_2 = \tilde{\Phi}_2(\xi_1, \xi_2) \right\} \quad (\text{by (37)}) \\
&= \left\{ w_1 = \tilde{\Phi}_1(\xi_1, \xi_2), w_2 = \tilde{\Phi}_2(\xi_1, \xi_2) \right\} \cap \left\{ v_1 = \Phi_1(\xi_1, v_2), v_2 = \Phi_2(\xi_2, v_1) \right\}
\end{aligned}$$

because $\left\{ v_1 = \Phi_1(\xi_1, v_2), v_2 = \Phi_2(\xi_2, v_1) \right\} = \Omega$ by (35c) and (35d)

$$\begin{aligned}
&= \left\{ w_1 = \tilde{\Phi}_1(\xi_1, \xi_2), w_2 = \tilde{\Phi}_2(\xi_1, \xi_2) \right\} \cap \left\{ v_1 = \tilde{\Phi}_1(\xi_1, \xi_2), v_2 = \tilde{\Phi}_2(\xi_1, \xi_2) \right\} \quad (\text{by (37)}) \\
&\subset \left\{ v_1 = w_1, v_2 = w_2 \right\}.
\end{aligned}$$

On the other hand, the reverse inclusion can be proved in the same way. Thus, we have obtained the equality (38).

• Second, we show that, for any subsets $\Theta'_1 \subset \Theta_1$ and $\Theta'_2 \subset \Theta_2$, and for any couple $(w_1, w_2) \in \Upsilon'_1 \times \Upsilon'_2$, we have that

$$\begin{aligned}
& \left\{ \theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2, v_1 = w_1, v_2 = w_2 \right\} \\
&= \left\{ \Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1 \right\} \cap \left\{ \Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_2(\xi_2, w_1) = w_2 \right\}. \quad (39)
\end{aligned}$$

Indeed, we have that

$$\begin{aligned}
& \left\{ \theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2, v_1 = w_1, v_2 = w_2 \right\} \\
&= \left\{ \Psi_1(\xi_1, v_2) \in \Theta'_1, \Psi_2(\xi_2, v_1) \in \Theta'_2, v_1 = w_1, v_2 = w_2 \right\} \quad (\text{by (35a) and (35b)}) \\
&= \left\{ \Psi_1(\xi_1, w_2) \in \Theta'_1, \Psi_2(\xi_2, w_1) \in \Theta'_2, v_1 = w_1, v_2 = w_2 \right\}
\end{aligned}$$

by substitution of the last two terms $v_1 = w_1$ and $v_2 = w_2$ in the first two terms

$$\begin{aligned}
&= \left\{ \Psi_1(\xi_1, w_2) \in \Theta'_1, \Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_1(\xi_1, v_2) = w_1, \Phi_2(\xi_2, v_1) = w_2, \right. \\
&\quad \left. v_1 = w_1, v_2 = w_2 \right\} \\
&\quad (\text{because } \left\{ \Phi_1(\xi_1, v_2) = v_1, \Phi_2(\xi_2, v_1) = v_2 \right\} = \Omega \text{ by (35c) and (35d)}) \\
&= \left\{ \Psi_1(\xi_1, w_2) \in \Theta'_1, \Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_1(\xi_1, w_2) = w_1, \Phi_2(\xi_2, w_1) = w_2, \right. \\
&\quad \left. v_1 = w_1, v_2 = w_2 \right\}
\end{aligned}$$

by substitution of the last two terms $v_1 = w_1$ and $v_2 = w_2$ in the two middle terms

$$= \left\{ \Psi_1(\xi_1, w_2) \in \Theta'_1, \Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_1(\xi_1, w_2) = w_1, \Phi_2(\xi_2, w_1) = w_2 \right\}.$$

because the system (36) of equations has a unique solution on $\Upsilon'_1 \times \Upsilon'_2$, so that $\Phi_1(\xi_1, w_2) = w_1$ and $\Phi_2(\xi_2, w_1) = w_2$ imply that $v_1 = w_1$ and $v_2 = w_2$ hold true by (35c) and (35d). Thus, we have obtained (39).

• Third, and finally, we show that the random variables θ_1 and θ_2 are independent when conditioned on (v_1, v_2) . For this purpose, we calculate, for any subsets $\Theta'_1 \subset \Theta_1$ and $\Theta'_2 \subset \Theta_2$, and for any couple $(w_1, w_2) \in \Upsilon'_1 \times \Upsilon'_2$:

$$\begin{aligned} & \mathbb{P}\left\{\theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2 \mid v_1 = w_1, v_2 = w_2\right\} \\ &= \frac{\mathbb{P}\left\{\theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2, v_1 = w_1, v_2 = w_2\right\}}{\mathbb{P}\left\{v_1 = w_1, v_2 = w_2\right\}} \end{aligned}$$

by definition of the conditional probability, and where all quantities are zero if the denominator is zero

$$\begin{aligned} &= \frac{\mathbb{P}\left\{\theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2, v_1 = w_1, v_2 = w_2\right\}}{\mathbb{P}\left\{\theta_1 \in \Theta_1, \theta_2 \in \Theta_2, v_1 = w_1, v_2 = w_2\right\}} \quad (\text{because } \left\{\theta_1 \in \Theta_1, \theta_2 \in \Theta_2\right\} = \Omega) \\ &= \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1\right\} \times \mathbb{P}\left\{\Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_2(\xi_2, w_1) = w_2\right\}}{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta_1, \Phi_1(\xi_1, w_2) = w_1\right\} \times \mathbb{P}\left\{\Psi_2(\xi_2, w_1) \in \Theta_2, \Phi_2(\xi_2, w_1) = w_2\right\}} \end{aligned}$$

by (39), and then using the assumption that the random variables ξ_1 and ξ_2 are independent

$$\begin{aligned} &= \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1\right\}}{\mathbb{P}\left\{\Phi_1(\xi_1, w_2) = w_1\right\}} \times \frac{\mathbb{P}\left\{\Psi_2(\xi_2, w_1) \in \Theta'_2, \Phi_2(\xi_2, w_1) = w_2\right\}}{\mathbb{P}\left\{\Phi_2(\xi_2, w_1) = w_2\right\}} \\ & \quad (\text{because } \left\{\Psi_1(\xi_1, w_2) \in \Theta_1, \Psi_2(\xi_2, w_1) \in \Theta_2\right\} = \Omega.) \end{aligned}$$

Then, we focus on the first term of the product and we write

$$\begin{aligned} & \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1\right\}}{\mathbb{P}\left\{\Phi_1(\xi_1, w_2) = w_1\right\}} \\ &= \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1\right\} \times \mathbb{P}\left\{\Phi_2(\xi_2, v_1) = w_2\right\}}{\mathbb{P}\left\{\Phi_1(\xi_1, w_2) = w_1\right\} \times \mathbb{P}\left\{\Phi_2(\xi_2, v_1) = w_2\right\}} \\ &= \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, \Phi_1(\xi_1, w_2) = w_1, \Phi_2(\xi_2, v_1) = w_2\right\}}{\mathbb{P}\left\{\Phi_1(\xi_1, w_2) = w_1, \Phi_2(\xi_2, v_1) = w_2\right\}} \end{aligned}$$

because the random variables ξ_1 and ξ_2 are independent

$$\begin{aligned} &= \frac{\mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1, v_1 = w_1, v_2 = w_2\right\}}{\mathbb{P}\left\{v_1 = w_1, v_2 = w_2\right\}} \quad (\text{by the equality (38)}) \\ &= \mathbb{P}\left\{\Psi_1(\xi_1, w_2) \in \Theta'_1 \mid v_1 = w_1, v_2 = w_2\right\} \quad (\text{by definition of the conditional probability}) \\ &= \mathbb{P}\left\{\theta_1 \in \Theta'_1 \mid v_1 = w_1, v_2 = w_2\right\}. \quad (\text{by (35a)}) \end{aligned}$$

Doing the same with the second term of the product, we get that

$$\begin{aligned} & \mathbb{P}\left\{\theta_1 \in \Theta'_1, \theta_2 \in \Theta'_2 \mid v_1 = w_1, v_2 = w_2\right\} \\ &= \mathbb{P}\left\{\theta_1 \in \Theta'_1 \mid v_1 = w_1, v_2 = w_2\right\} \times \mathbb{P}\left\{\theta_2 \in \Theta'_2 \mid v_1 = w_1, v_2 = w_2\right\}. \end{aligned}$$

This ends the proof. □

5.3.2 Graphical discussion on Lemma 22

A graphical representation of the system of random variables described in Lemma 22 necessarily contains a cycle between v_1 and v_2 , because of (35c)–(35d). As a consequence, classical results cannot be applied.

By contrast, using the reparametrization (37) of Equations (35c) and (35d) — giving $v_1 = \tilde{\Phi}_1(\xi_1, \xi_2)$, and $v_2 = \tilde{\Phi}_2(\xi_1, \xi_2)$ — we obtain a graphical representation which is free of cycle. However, this is at the cost of losing some properties of the initial parametrization. Indeed, system (35) becomes

$$\theta_1 = \Psi_1(\xi_1, v_2), \tag{42a}$$

$$\theta_2 = \Psi_2(\xi_2, v_1), \tag{42b}$$

$$v_1 = \tilde{\Phi}_1(\xi_1, \xi_2), \tag{42c}$$

$$v_2 = \tilde{\Phi}_2(\xi_1, \xi_2), \tag{42d}$$

and its DAG representation is now the one displayed in Figure 6. In Figure 6, we observe that there exists an unblocked path $\theta_1 \leftarrow \xi_1 \rightarrow v_1 \leftarrow \xi_2 \rightarrow \theta_2$ from θ_1 to θ_2 . As a consequence, we cannot conclude about the conditional independence of θ_1 and θ_2 with respect to (v_1, v_2) .

By contrast, with Lemma 22 we reach the conclusion that the random variables θ_1 and θ_2 are independent when conditioned on (v_1, v_2) .

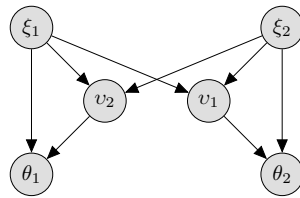


Figure 6: DAG representation of the system of equations (42)

5.4 Discrete or continuous? It does matter

It is notable that Lemma 22 seems to be in contradiction with an example from [21] (recently cited in [4, Example 6.1]).

Example 23 (from [21]) *Sirtes considers the following model (with R_X, R_Y, R_Z, R_W being standard independent normal random variables):*

$$X = R_X \tag{43a}$$

$$Y = R_Y \tag{43b}$$

$$Z = WY + R_Z \tag{43c}$$

$$W = ZX + R_W \tag{43d}$$

Sirtes shows that X and Y are not independent given (Z, W) . However if we set (with obvious notations related to Lemma 22)

$$v_1 = Z, \tag{44a}$$

$$v_2 = W, \tag{44b}$$

$$\xi_1 = (R_Z, Y), \tag{44c}$$

$$\xi_2 = (R_W, X), \tag{44d}$$

$$\theta_1 = \psi_1(\xi_1, v_2) = \psi_1((R_Z, Y), v_2) := Y, \tag{44e}$$

$$\theta_2 = \psi_2(\xi_2, v_1) = \psi_2((R_W, X), v_1) := X, \tag{44f}$$

$$\Phi_1(\xi_1, v_2) = \Phi_1((R_Z, Y), W) := WY + R_Z, \tag{44g}$$

$$\Phi_2(\xi_2, v_1) = \Phi_2((R_W, X), Z) := ZX + R_W, \tag{44h}$$

then we see that a countable version of this example could be treated with Lemma 22. In particular, X and Y are independent given (Z, W) , which is different from Sirtes's conclusion.

The countable assumption in Lemma 22 seems to draw a line between the systems described in the present paper and the approach presented in [4]. Hence, we have an example of system for which a conditional independence property depends on whether the codomain of ξ_1 and ξ_2 is discrete or continuous. We mention that a phenomenon of the same flavour is discussed in [3].

5.5 Topological separation implies conditional independence

We now use the results obtained in §5.2 and in §5.3 to state a general result of conditional independence (Theorem 24), a corollary of which (Theorem 25) constitutes a new version of Pearl's rule of do-calculus.

Theorem 24 *We suppose that the assumptions of Lemma 20 are satisfied. Moreover, we suppose that the set Ω in (24) is equipped with a probability $\mathbb{P} = \bigotimes_{a \in \mathbb{A}} \mathbb{P}_a$ where each \mathbb{P}_a is a probability on $(\Omega_a, \mathcal{F}_a)$.*

We define the following pushforward probability \mathbb{Q}_λ on $(\mathbb{H}, \mathcal{H})$, in (10), by

$$\mathbb{Q}_\lambda = \mathbb{P} \circ S_\lambda^{-1}. \tag{45}$$

Then, $(\mathbb{H}, \mathcal{H}, \mathbb{Q}_\lambda)$ is a probability space, and the two projections $\pi_{\overline{Y}^{W,H}} : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_{\overline{Y}^{W,H}}, \mathcal{U}_{\overline{Y}^{W,H}})$ and $\pi_{\overline{Z}^{W,H}} : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_{\overline{Z}^{W,H}}, \mathcal{U}_{\overline{Z}^{W,H}})$ as in (25c) are independent under \mathbb{Q}_λ , conditionally on the subset $H \subset \mathbb{H}$ and on the projection $\pi_W : (\mathbb{H}, \mathcal{H}) \rightarrow (\mathbb{U}_W, \mathcal{U}_W)$.

Theorem 24 claims that, for any values $u_Y \in \mathbb{U}_{\bar{Y}^{w,H}}$, $u_Z \in \mathbb{U}_{\bar{Z}^{w,H}}$ and $u_W \in \mathbb{U}_W$, we have that

$$\begin{aligned} & \mathbb{Q}_\lambda \left(\pi_{\bar{Y}^{w,H}}(h) = u_Y, \pi_{\bar{Z}^{w,H}}(h) = u_Z \mid h \in H, \pi_W(h) = u_W \right) \\ &= \mathbb{Q}_\lambda \left(\pi_{\bar{Y}^{w,H}}(h) = u_Y \mid h \in H, \pi_W(h) = u_W \right) \\ & \quad \times \mathbb{Q}_\lambda \left(\pi_{\bar{Z}^{w,H}}(h) = u_Z \mid h \in H, \pi_W(h) = u_W \right). \end{aligned}$$

Proof. If $\mathbb{P}(S_\lambda^{-1}(H)) = 0$, conditional independence is trivial (and meaningless!). We suppose that $\mathbb{P}(S_\lambda^{-1}(H)) > 0$ and we instantiate Lemma 22 with

- probability space $\tilde{\Omega} = S_\lambda^{-1}(H)$ with renormalized probability $\tilde{\mathbb{P}} = \mathbb{P}/\mathbb{P}(S_\lambda^{-1}(H))$,
- six measurable spaces $\Xi_1 = \Omega_{\bar{Y}}$, $\Xi_2 = \Omega_{\bar{Z}}$, $\Upsilon_1 = \mathbb{U}_{W_Y}$, $\Upsilon_2 = \mathbb{U}_{W_Z}$, $\Theta_1 = \mathbb{U}_{Y \cup Y'}$, $\Theta_2 = \mathbb{U}_{Z \cup Z'}$,
- four measurable mappings $\Psi_1 = \pi_{Y \cup Y'} \circ \tilde{M}_{\lambda_Y}$, $\Psi_2 = \pi_{Z \cup Z'} \circ \tilde{M}_{\lambda_Z}$, $\Phi_1 = \pi_{W_Y} \circ \tilde{M}_{\lambda_Y}$, $\Phi_2 = \pi_{W_Z} \circ \tilde{M}_{\lambda_Z}$,
- six random variables $\xi_1(\omega) = \omega_{\bar{Y}}$, $\xi_2(\omega) = \omega_{\bar{Z}}$, for all $\omega \in \tilde{\Omega}$, and $\theta_1 = \pi_{Y \cup Y'} \circ S_\lambda$, $\theta_2 = \pi_{Z \cup Z'} \circ S_\lambda$, $v_1 = \pi_{W_Y} \circ S_\lambda$, $v_2 = \pi_{W_Z} \circ S_\lambda$ on $\tilde{\Omega}$.

By assumption, the set Ω in (24) is equipped with a probability $\mathbb{P} = \bigotimes_{a \in \mathbb{A}} \mathbb{P}_a$ where each \mathbb{P}_a is a probability on $(\Omega_a, \mathcal{F}_a)$. Because of the product structure, the random variables ξ_1 and ξ_2 are independent with respect to $\tilde{\mathbb{P}}$.

As the assumptions of Lemma 20 are satisfied, Equation (28d) holds true, that is, we have that

$$M_\lambda(\omega) = \left(\tilde{M}_{\lambda_{\bar{Y}}} \left(\omega_{\bar{Y}}, \pi_{W_Z}(S_\lambda(\omega)) \right), \tilde{M}_{\lambda_{\bar{Z}}} \left(\omega_{\bar{Z}}, \pi_{W_Y}(S_\lambda(\omega)) \right), \tilde{M}_{\lambda_E} \left(\omega_E, \pi_{\bar{Y} \cup \bar{Z}}(S_\lambda(\omega)) \right) \right), \quad \forall \omega \in S_\lambda^{-1}(H).$$

Thus, the assumptions of Lemma 22 are satisfied, and we conclude that the random variables θ_1 and θ_2 are independent under the probability $\tilde{\mathbb{P}}$, when conditioned on (v_1, v_2) .

In other words, we have obtained that $\pi_{Y \cup Y'} \circ S_\lambda = \pi_{\bar{Y}} \circ S_\lambda$ and $\pi_{Z \cup Z'} \circ S_\lambda = \pi_{\bar{Z}} \circ S_\lambda$ are independent random variables, when conditioned on $\pi_{W_Y} \circ S_\lambda$ and $\pi_{W_Z} \circ S_\lambda$ under the probability $\tilde{\mathbb{P}}$. We deduce that $\pi_{\bar{Y}}$ and $\pi_{\bar{Z}}$ are independent when conditioned on π_{W_Y} and π_{W_Z} under the probability $\mathbb{Q}_\lambda = \mathbb{P} \circ S_\lambda^{-1}$.

This ends the proof. \square

5.6 Topological separation implies the do-calculus

Next we deduce from Theorem 24 a variant of Pearl's do-calculus.

Theorem 25 (Do-calculus in W-models) *Under the assumptions of Theorem 24, the projection π_Y has the same conditional distribution under \mathbb{Q}_λ , whether the conditioning is w.r.t. the subset $H \subset \mathbb{H}$, the projection π_W and the projection $\pi_{\bar{Z}^{w,H}}$, or is only w.r.t. the subset $H \subset \mathbb{H}$ and the projection π_W .*

Proof. By Theorem 24, when conditioning with respect to H and π_W , π_Y and $\pi_{\bar{Z}^{W,H}}$ are independent. This implies (see for example [24, Proposition 2.4 (c)]) in particular that $\pi_{\bar{Z}^{W,H}}$ can be removed from the conditioning above mentioned. \square

We have proved, loosely speaking, that

$$Y \perp\!\!\!\perp_t Z \mid (W, H) \implies \mathbb{Q}_\lambda(h_Y | h_W, h_{\bar{Z}^{W,H}}, H) = \mathbb{Q}_\lambda(h_Y | h_W, H). \quad (46)$$

In particular

$$Y \perp\!\!\!\perp_t Z \mid (W, H) \implies \mathbb{Q}_\lambda(h_Y | h_W, h_Z, H) = \mathbb{Q}_\lambda(h_Y | h_W, H). \quad (47)$$

We stress the conciseness of Theorem 25 — permitted by the notions introduced in this paper — as we now show that it implies the three rules of Pearl, as well as the following two recent results. As already mentioned in Example 7, the authors in [23] manage to summarize the three rules of do-calculus thanks to the notion of context specific independence. They rely on so-called *labeled DAG* that can be turned into a context specific DAG by removing the arcs that are deactivated (spurious) in the context of interest. In the formalism that we propose, such context is represented by a subset of \mathbb{H} . Indeed, if we denote by $H \in \mathbb{H}$ the context for which an arc (a, b) is deactivated (in the language of [23]), we represent this by the following two properties: $a \notin \mathcal{P}_{\emptyset, H} b$, $a \in \mathcal{P}_{\emptyset, H^c} b$. Such a property can be also be encoded in the information set of agent b . As a consequence, there is a mapping from the model introduced in [23] to W-models.

To introduce the next result, we will allow some abuse of notations to make our notations as close as possible to the literature we are comparing with. We will use, for $B \subset \mathbb{A}$ and $u_B \in \mathbb{U}_B$, the notation $[h_B = u_B] = \{h \in \mathbb{H} \mid h_B = u_B\}$. Then, Rule 1 in [23] rewrites, in our setting, as

$$Y \perp\!\!\!\perp_{t(X, h_{\bar{X}} = u_{\bar{X}})} Z \implies \mathbb{Q}(h_Y | h_Z, h_X, h_{\bar{X}} = u_{\bar{X}}) = \mathbb{Q}(h_Y | h_X, h_{\bar{X}} = u_{\bar{X}}) \quad (48)$$

where $X, \bar{X} \subset \mathbb{A}$ and for a given value $u_{\bar{x}}$.

Proposition 26 *Rule 1 from [23] can be deduced from Theorem 25. In particular, Theorem 25 subsumes Pearl's do-calculus from [15].*

Proof. If we set $W = X$ and $H = \{h \in \mathbb{H}; h_{\bar{X}} = u_{\bar{X}}\}$ in Equation (47) (obtained with Theorem 25) we obtain (48) which is Rule 1 from [23]. The proof of Theorem 2 from [23] states that this rule implies in particular the rules of Pearl's do-calculus. \square

6 Discussion

In this paper, we simplify and generalize the do-calculus by leveraging the concept of information field, using Witsenhausen’s intrinsic model. The do-calculus is reduced to one rule. We underline that the results are consequences of the information structure, but have nothing to do with the probability. For most cases, one only needs to understand the notion of inverse image to work with information fields on top of SCMs and DAGs. In exchange, information fields provide a compact, unifying and versatile language that brings new intuitions on the causal structure of the problem.

For instance, we have illustrated why the notion of topological separation is practical: once the splitting of the conditioning variables known, checking that an intersection is empty is easier than checking a blocking condition on a collection of paths. We prove in [9] that the topological separation is equivalent to the d-separation on DAGs.

The Information Dependency Model is a good candidate to bring uniformity and consistency in lieu of *ad hoc* frameworks. It can be a temporary detour to introduce new notions, for instance the Definition 4 of conditional precedence would have been harder to express with the SCM as primitive.

In addition, we have presented and solved an example that cannot be handled easily with the current state of the literature.

Last, we mention that the notion of well-posedness we use was introduced in [27] half a century ago for another field of applied mathematics. It is interesting to observe that this notion could serve a new purpose in the field of causal inference.

Further work includes drawing connections with other research programs, such as questions related to identification causal structure [19, 20, 22] or extensions of do-calculus [7]. As argued in Sect. 5.4, there is a fundamental difference between the discrete and continuous case that calls for different tooling; in this regard, it would be interesting to study the connections of this work with [4, 12].

Aknowledgements

We thank the organizers of the Causal Discovery and Causality-Inspired Machine Learning Workshop at Neural Information Processing Systems, where we could present this work on 11 December 2020. We thank Sridhar Mahadevan for interesting exchanges.

References

- [1] M. S. Andersland and D. Teneketzis. Information structures, causality and nonsequential stochastic control I: Design-independent properties. *SIAM J. Control Optim.*, 30(6):1447–1475, 1992.

- [2] M. S. Andersland and D. Teneketzis. Information structures, causality, and nonsequential stochastic control. II: Design-dependent properties. *SIAM Journal on Control and Optimization*, 32(6):1726–1751, 1994.
- [3] M. Barbie and A. Gupta. The topology of information on the space of probability measures over Polish spaces. *Journal of Mathematical Economics*, 52:98–111, 2014.
- [4] S. Bongers, P. Forré, J. Peters, B. Schölkopf, and J. M. Mooij. Foundations of structural causal models with cycles and latent variables. *arXiv preprint arXiv:1611.06221*, 2020.
- [5] S. Bongers and J. M. Mooij. From random differential equations to structural causal models: The stochastic case. *arXiv preprint arXiv:1803.08784*, 2018.
- [6] J.-P. Chancelier, M. De Lara, and B. Heymann. Conditional separation as a binary relation, 2021. Preprint.
- [7] J. Correa and E. Bareinboim. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *AAAI 2020 : The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.
- [8] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media, 2006.
- [9] M. De Lara, J.-P. Chancelier, and B. Heymann. Topological conditional separation, 2021. Preprint.
- [10] C. Dellacherie and P. A. Meyer. *Probabilités et potentiel*. Hermann, Paris, 1975.
- [11] J. L. Doob. *Stochastic processes*, volume 10. New York Wiley, 1953.
- [12] P. Forré and J. M. Mooij. Causal calculus in the presence of cycles, latent confounders and selection bias. In R. P. Adams and V. Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 71–80. PMLR, Tel Aviv, Israel, 22–25 Jul 2020.
- [13] J. Y. Halpern. Axiomatizing causal reasoning. *J. Artificial Intelligence Res.*, 12:317–337, 2000.
- [14] F. Lattimore and D. Rohde. Replacing the do-calculus with Bayes rule. *arXiv preprint arXiv:1906.07125*, 2019.
- [15] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- [16] J. Pearl. *Causality: Models, reasoning, and inference*. Cambridge University Press, second edition edition, 2011.

- [17] J. Pearl and D. Mackenzie. *The book of Why: the new science of cause and effect*. Basic Books, 2018.
- [18] J. Peters, D. Janzing, and B. Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.
- [19] I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06*, volume 2, pages 1219–1226. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.
- [20] I. Shpitser and J. Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9(Sep):1941–1979, 2008.
- [21] P. Spirtes. Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI’95*, pages 491–498, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
- [22] S. Tikka, A. Hyttinen, and J. Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *arXiv preprint arXiv:1902.01073*, 2019.
- [23] S. Tikka, A. Hyttinen, and J. Karvanen. Identifying causal effects via context-specific independence relations. In *Advances in Neural Information Processing Systems*, pages 2804–2814, 2019.
- [24] C. van Putten and J. H. van Schuppen. Invariance properties of the conditional independence relation. *Ann. Probab.*, 13(3):934–945, 1985.
- [25] J. Winn. Causality with gates. In *Artificial Intelligence and Statistics*, pages 1314–1322, 2012.
- [26] H. S. Witsenhausen. On information structures, feedback and causality. *SIAM J. Control*, 9(2):149–160, May 1971.
- [27] H. S. Witsenhausen. The intrinsic model for discrete stochastic control: Some open problems. In A. Bensoussan and J. L. Lions, editors, *Control Theory, Numerical Methods and Computer Systems Modelling*, volume 107 of *Lecture Notes in Economics and Mathematical Systems*, pages 322–335. Springer-Verlag, 1975.