Studies in Logic, Vol. 14, No. 6 (2021): 1–24
PII: 1674-3202(2021)-06-0001-24

Markov Categories, Causal Theories, and the Do-calculus*

Yimu Yin Jiji Zhang

Abstract. We give a category-theoretic treatment of causal models that formalizes the syntax for causal reasoning over a directed acyclic graph (DAG) by associating a free Markov category with the DAG in a canonical way. This framework enables us to define and study important concepts in causal reasoning from an abstract and "purely causal" point of view, such as causal independence/separation, causal conditionals, and decomposition of intervention effects. Our results regarding these concepts abstract away from the details of the commonly adopted causal models such as (recursive) structural equation models or causal Bayesian networks. They are therefore more widely applicable and in a way conceptually clearer. Our results are also intimately related to Judea Pearl's celebrated do-calculus, and yield a syntactic version of a core part of the calculus that is inherited in all causal models. In particular, it induces a simpler and specialized version of Pearl's do-calculus in the context of causal Bayesian networks, which we show is as strong as the full version.

1 Introduction

Causal models based on directed acyclic graphs (DAGs), such as recursive structural equation models ([3, 4]) and causal Bayesian networks ([13, 11]), have been vigorously studied and widely applied as powerful tools for causal reasoning. However, from a logical point of view, the syntax underlying such causal models is usually left implicit or even obscure in the literature. This lacuna is fixed in recent category-theoretic work on the subject ([1, 6]), where the distinction between syntax and semantics is made clear in the style of F.W. Lawvere's functorial semantics ([7]). Specifically, the syntax is provided by a monoidal category of a certain kind induced by a DAG, and a distinguished class of morphisms therein can be viewed as *syntactic* causal effects, which may then be interpreted in various ways. Causal Bayesian networks, for example, interpret a causal effect of this kind with a stochastic matrix that represents probability distributions over the outcome-variables resulting

Received 2021-08-31

Yimu Yin Pasadena, California
yimu.yin@hotmail.com
Jiji Zhang Department of Religion and Philosophy, Hong Kong Baptist University
zhangjiji@hkbu.edu.hk

*This research was supported by the Research Grants Council of Hong Kong under the General Research Fund (13602818). Zhang would also like to acknowledge the support by a start-up grant from the Hong Kong Baptist University.

from interventions on the treatment-variables. In this light, a causal Bayesian network is a functor—a structure-preserving mapping—from the syntax category to a category whose morphisms are stochastic matrices. As another example (and a degenerate case of causal Bayesian networks), deterministic structural equation models interpret a causal effect of this kind as a function that represents how the values of the outcome-variables depend on those of the treatment-variables. Thus, a (recursive) deterministic structural equation model may be viewed as a functor from the syntax category to a category whose morphisms are functions.

In this paper we build on this category-theoretic framework and study some important concepts in DAG-based causal reasoning from a syntactic and more abstract perspective. In particular, we work with the categories defined in [1], called *causal theories*, with an extra constraint to make them Markov categories in the sense of [2]. We study the morphisms in that category that correspond to what we call syntactic causal effects, using the graphical language of string diagrams as vehicles for our arguments. One of our main results concerns the decomposition or disintegration of causal effect morphisms, or in the terminology of [2], the existence of a conditional for a causal effect morphism. Roughly, this refers to the property that the causal effect of x on y and z can be decomposed into the causal effect of x on y followed by that of x and y on z. We derive a condition that is sufficient and necessary for decomposability in a causal theory. Interestingly, the condition is precisely the condition in a specialized version of the second rule of Judea Pearl's do-calculus ([11]). This agreement is of course not a coincidence and has, we submit, several instructive implications. The other rules of the do-calculus have more straightforward counterparts in terms of causal effect morphisms, and the upshot is a generic do-calculus at the syntactic level.

This generic calculus, we argue, captures the "causal core" of reasoning about interventions, and is automatically inherited in all causal models, including but not limited to the popular probabilistic ones. In particular, it induces a simpler and specialized version of Pearl's do-calculus in the context of causal Bayesian networks. Importantly, we show that given the probability calculus, the simpler and specialized version entails the full version of the do-calculus, corroborating our contention that the generic do-calculus corresponds to the purely causal component of the well-known probabilistic do-calculus.

The rest of the paper is organized as follows. In Section 2, we introduce the basics of category theory and the intuitive language of string diagrams, leading to the notion of a Markov category. In Section 3, we define causal theories as an abstraction of causal Bayesian networks and as free Markov categories, and highlight a class of morphisms constructed in [1], which we call "causal effect morphisms". Section 4 presents some results about causal effect morphisms, which yield a more abstract and syntactic counterpart to Pearl's do-calculus. We show in Section 5 that the syntactic

do-calculus entails a simpler and specialized version of probabilistic do-calculus, and that, despite its simplicity, the specialized version is actually as strong as the full version.

2 Markov Categories

For the sake of space and readability, we will only describe the notions of category theory that are essential for understanding this paper, and introduce the axioms for a Markov category using the language of string diagrams. For readers interested in learning more about category theory and string diagrams, we recommend the canonical treatises [8] and [12], among other excellent textbooks and surveys.

A category C consists of two types of entities: *objects* A, B, C, ... and *arrows* f, g, h, ..., subject to the following three rules:

- For each arrow f there are given two objects $\mathrm{dom}(f)$ and $\mathrm{cod}(f)$, called the *domain* and the *codomain* of f. We usually write $f : A \longrightarrow B$ to indicate that $A = \mathrm{dom}(f)$ and $B = \mathrm{cod}(f)$.
- Given two arrows $f : A \longrightarrow B$ and $g : B \longrightarrow C$, that is, $\mathrm{cod}(f) = \mathrm{dom}(g)$, there is a third arrow $g \circ f : A \longrightarrow C$, called the *composition* of f and g.
- For each object A there is an arrow $1_A : A \longrightarrow A$, called the *identity* or *unit* arrow of A.

In addition, the obvious *unitality* and *associativity* laws hold for compositions: for all $f : A \longrightarrow B$, $g : B \longrightarrow C$, and $h : C \longrightarrow D$,

$$1_B \circ f = f, \quad f \circ 1_A = f, \quad (h \circ g) \circ f = h \circ (g \circ f).$$

An arrow in category theory is also called a *morphism* or a *map*. Here is a more formal definition:

Definition 2.1. Let C be a quadruple $(\mathrm{C}_0, \mathrm{C}_1, \mathrm{dom}, \mathrm{cod})$, where C_0 is referred to as a class of *objects*, C_1 is referred to as a class of *morphisms*, and $\mathrm{dom} : \mathrm{C}_1 \longrightarrow \mathrm{C}_0$, $\mathrm{cod} : \mathrm{C}_1 \longrightarrow \mathrm{C}_0$ are functions. A morphism f in C_1 is usually written as $f : A \longrightarrow B$ with $\mathrm{dom}(f) = A$ and $\mathrm{cod}(f) = B$. For each pair of objects A, B in C_0, the class of all morphisms f with $\mathrm{dom}(f) = A$ and $\mathrm{cod}(f) = B$ is denoted by $\mathrm{hom}_{\mathrm{C}}(A, B)$.

Let $\mathrm{C}_2 = \{(f, g) \in \mathrm{C}_1 \times \mathrm{C}_1 \mid \mathrm{cod}(f) = \mathrm{dom}(g)\}$. We say that C is a *category* if it also comes with a morphism $1_A : A \longrightarrow A$ for every $A \in \mathrm{C}_0$, called the *identity morphism* of A, and a function $\circ : \mathrm{C}_2 \longrightarrow \mathrm{C}_1$, called *composition*, subject to the associativity and unitality laws given above.

Often we just write $x \in \mathrm{C}$ and let the context determine whether x is an object or a morphism.

A paradigmatic example of a category is the category Set, containing sets as objects and functions as morphisms. In this category, the composition of morphisms

is just the composition of functions and for each object A, the identity morphism is just the identity function.

It is helpful to think of a morphism as an abstract function, or a box with input wires and output wires, as in the graphical language of string diagrams. The four rudiments of a category are depicted in such a graphical language as follows:

object A | morphism $f : A \longrightarrow B$ | identity 1_A | composition $g \circ f$

(2.1)

Note that string diagrams are parsed in the lower-left to upper-right order.

Remark 2.2. A string diagram is a topological graph in which every edge is labelled with an object and every vertex with a morphism. ([12]) Object labels such as A, B are usually omitted except when they are needed for clarity or emphasis. A labelled vertex is also called a *node*, and is often drawn as a box such as $\boxed{f}$ for readability. Just as in the usual symbolic formalism, a morphism f may be represented by many string diagrams.

Categories may serve as objects in a "higher" category, and the morphisms between categories are known as functors:

Definition 2.3. Let C, D be categories and F a pair of mappings $F_0 : \mathrm{C}_0 \longrightarrow \mathrm{D}_0$, $F_1 : \mathrm{C}_1 \longrightarrow \mathrm{D}_1$. Then F is a *functor*, written as $F : \mathrm{C} \longrightarrow \mathrm{D}$, if the following three conditions, corresponding to the three conditions for a category, are satisfied:

- F preserves domains and codomains, that is, $F_1(f : A \longrightarrow B)$ is a morphism $F_0(A) \longrightarrow F_0(B)$ for all morphisms $f \in \mathrm{C}_1$.
- F preserves compositions, that is, $F_1(g \circ f) = F_1(g) \circ F_1(f)$ for all compositions $g \circ f \in \mathrm{C}_1$.
- F preserves identities, that is, $F_1(1_A) = 1_{F_0(A)}$ for all objects $A \in \mathrm{C}_0$.

Compositions of functors may be defined using composition of mappings. Then it is routine to check that categories and functors form a "higher" category.

We now introduce more structures needed for our purpose. Start with the (strict) monoidal structure:

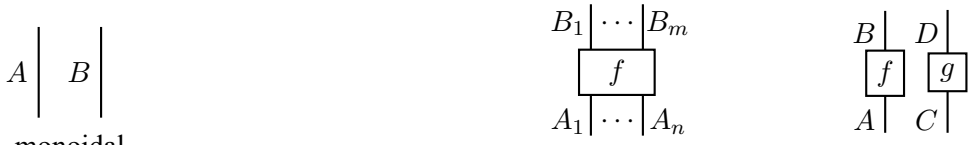
Definition 2.4. A *strict monoidal category* is a category C equipped with a functor $\otimes : \mathrm{C} \times \mathrm{C} \longrightarrow \mathrm{C}$, called the *monoidal product*, and a distinguished object $\mathbf{I}$, called

the *monoidal unit* (of the monoidal product), that satisfy associativity and unitality:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C), \quad \mathbf{I} \otimes A = A = A \otimes \mathbf{I} \tag{2.2}$$

Many commonly encountered monoidal categories are actually not strict because equation (2.2) holds only "up to isomorphism". For example, the category Set has an obvious monoidal product, which is just the Cartesian product (of sets and of functions). The monoidal unit is any singleton set, but unitality is a matter of isomorphism rather than strict identity. The formal definition of a (possibly non-strict) monoidal category is rather more complex and requires the notion of a natural transformation, which we omit to keep things simple. The monoidal categories we will focus on in this paper are strict.[1]

Since the monoidal product is a functor, it applies to both objects and morphisms in the category. Thus the graphical syntax in (2.1) is extended for monoidal categories as follows:

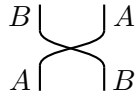


monoidal product of objects | monoidal unit (empty diagram) | morphism $f : \bigotimes_i A_i \longrightarrow \bigotimes_j B_j$ | monoidal product of morphisms (2.3)

Notation 2.5. We denote by A^n the monoidal product of n copies of an object A itself; this includes the empty product $A^0 = \mathbf{I}$. When an object of the form $A_1 \otimes \ldots \otimes A_k$ is introduced in the discussion, the indices are in general meant to indicate the ordering in which the monoidal product is taken.

For the present purpose, the monoidal structure is especially useful because it can be used to express causal processes or mechanisms that run in parallel, as is visualized in (2.3), whereas composition is used to express those that run in sequence.

A *symmetric* monoidal category is a monoidal category with natural isomorphisms $\sigma_{AB} : A \otimes B \cong B \otimes A$ that satisfy certain coherence conditions (the details do not matter for the present purpose). Graphically, a symmetry isomorphism is depicted as a crossing:

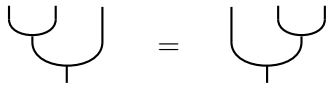


symmetry σ_{AB} (2.4)

[1] By S. Mac Lane's coherence theorem ([8, Theorem XI 3.1]), every monoidal category is monoidally equivalent to a strict monoidal category. So there is a sense in which we can treat monoidal categories as if they are all strict (even though they are not). See [12] for more on how this is justified.

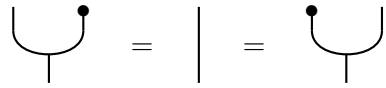
Finally, we can define a Markov category, following the lead of [2].

Definition 2.6. A *Markov category* is a symmetric monoidal category such that for each object A in C there are distinguished morphisms $\delta_A : A \longrightarrow A \otimes A$, called the *duplicate* on A, and $\epsilon_A : A \longrightarrow \mathbf{I}$, called the *discard* on A, that satisfy the following:



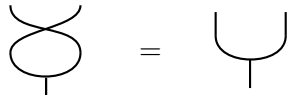
coassociativity

$(\delta_A \otimes 1_A) \circ \delta_A = (1_A \otimes \delta_A) \circ \delta_A$



counitality

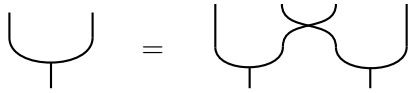
$(1_A \otimes \epsilon_A) \circ \delta_A = 1_A = (\epsilon_A \otimes 1_A) \circ \delta_n$



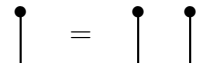
cocommutativity

$\sigma_{AA} \circ \delta_A = \delta_A$

(2.5)

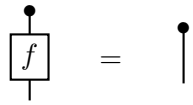


$\delta_{A\otimes B} = (1_A \otimes \sigma_{BA} \otimes 1_B) \circ (\delta_A \otimes \delta_B)$



$\epsilon_{A\otimes B} = \epsilon_A \otimes \epsilon_B$

(2.6)



$\epsilon_B \circ f = \epsilon_A$

(2.7)

Notation 2.7. Recall that our convention is to draw a string diagram in the lower-left to upper-right direction. So, above, the duplicate $\delta_A : A \longrightarrow A \otimes A$ is depicted as an upward fork ⋃ and the discard $\epsilon_A : A \longrightarrow \mathbf{I}$ an upward dead-end ┬.

Thus a Markov category is endowed with both a symmetric monoidal structure and additional duplicate morphisms and discard morphisms that satisfy (2.5)-(2.7). For our purpose, duplicate morphisms are needed mainly to express the same input entering several causal processes, and discard morphisms are needed to express ignoring or marginalizing over some outcomes of a causal process. The equations in (2.5) are axioms for the so-called *cocommutative monoidal comonoid*, and the equations in (2.6) express the condition that duplicates and discards respect the monoidal product. All these axioms are fairly intuitive. The equation in (2.7) says roughly that

discarding the output of a morphism is identical to discarding the input in the first place. This condition is not explicitly imposed in [1] but is actually needed for a main result therein (more on this later).

For categories with additional structures, the notion of a functor can be strengthened accordingly. For example, a *(strong) Markov functor* between two Markov categories is a functor that preserves the relevant structures (see [2, Definition 10.14] for details.)

3 Causal Theories and Effect Morphisms

In this section we describe the central object of study in this paper, a syntax category for DAG-based causal models defined in [1], called a causal theory. The framework is an abstraction of causal Bayesian networks, so we first review the latter in 3.1, and then introduce causal theories as free Markov categories in 3.2. In 3.3 we highlight a class of morphisms constructed in [1], which we refer to as causal effect morphisms. Our main results are concerned with these morphisms.

3.1 Causal Bayesian networks

A *directed graph* is a quadruple $G = (V, A, s, t)$, where V, A are sets and $s : A \longrightarrow V$, $t : A \longrightarrow V$ are functions. Elements in V are called *vertices* of G and those in A are *directed edges* or *arrows* of G. For $a \in A$, $s(a)$ is the *source* of a and $t(a)$ the *target* of a. G is *finite* if both V and A are. It is *simple* if, for all $a, a' \in A$, we have $a = a'$ whenever $s(a) = s(a')$ and $t(a) = t(a')$. We consider only finite simple graphs in this paper. A sequence of distinct arrows $a_1, \ldots, a_n \in A(G)$ is a *directed path*, starting from $s(a_1)$ and ending at $t(a_n)$, if $s(a_{i+1}) = t(a_i)$ for $1 \leq i < n$. It is a *cycle* if, in addition, $s(a_1) = t(a_n)$. G is *acyclic* if it contains no cycle. For $x, y \in V(G)$, x is called a *parent* of y and y a *child* of x if for some $a \in A(G)$, $s(a) = x$ and $t(a) = y$.

A *Bayesian network* (BN) over a set of (categorical) random variables $\mathbf{V}$ consists of a triple (G, P, υ), where G is a directed acyclic graph (DAG), P is a joint probability law of $\mathbf{V}$, and $\upsilon : V(G) \to \mathbf{V}$ is a bijection between the vertices of G and the random variables. Following common practice, we will leave the bijection υ implicit and simply identify $V(G)$ with $\mathbf{V}$, and call G a DAG over $\mathbf{V}$. The defining condition of a BN is that G and P satisfy a Markov condition, which requires that P can be factorized according to G as follows:

$$P(\mathbf{V}) = \prod_{X \in \mathbf{V}} P(X | \operatorname{pa}_G(X)), \tag{3.1}$$

where $\operatorname{pa}_G(X)$ denotes the set of parents of X in G. When G is sufficiently sparse, the factorization enables efficient computations of various probabilities entailed by

the joint probability law, which makes the BN useful for probabilistic reasoning. ([9])

The DAG in a BN usually lends itself to a causal interpretation, as a representation of the qualitative causal structure. With this causal reading, the BN framework can be extended to handle reasoning about effects of interventions. ([13, 11]) Specifically, a *causal Bayesian network* (CBN) over $\mathbf{V}$ does not represent just one joint probability law, but a number of *interventional* probability distributions. The standard setup is that for every subset $\mathbf{T} \subseteq \mathbf{V}$ and every possible value configuration $\mathbf{t}$ for $\mathbf{T}$, there is a probability distribution resulting from an (exogenous) intervention that forces $\mathbf{T}$ to take value $\mathbf{t}$. Such an interventional distribution, denoted as $P(\mathbf{V}|do(\mathbf{T} = \mathbf{t}))$ using Pearl's ([11]) do-operator, is assumed to be equal to a truncated factorization:

$$P(\mathbf{V}|do(\mathbf{T} = \mathbf{t})) = \begin{cases} \prod_{X \in \mathbf{V} \setminus \mathbf{T}} P(X|\,\mathrm{pa}_G(X)) & \text{for values of } \mathbf{V} \text{ consistent with } \mathbf{T} = \mathbf{t}, \\ 0 & \text{for values of } \mathbf{V} \text{ inconsistent with } \mathbf{T} = \mathbf{t}. \end{cases} \tag{3.2}$$

As a special case, when $\mathbf{T} = \emptyset$, we obtain the factorization (3.1) of the *pre-intervention* distribution. Equation (3.2) can be viewed as the defining axiom for the CBN, sometimes referred to as the *intervention principle*. ([17])

Note two key ideas in this formulation of a CBN: (1) for each $X \in \mathbf{V}$, $P(X|\,\mathrm{pa}_G(X))$ encodes a *modular* causal process or mechanism (when $\mathrm{pa}_G(X) = \emptyset$, $P(X)$ is taken to encode an *exogenous* mechanism for X), and the whole causal system is composed of these causal modules; (2) an intervention breaks the modules for its target variables but does not affect the other modules (hence the truncated factorization). Put this way, $P(X|\,\mathrm{pa}_G(X))$ is a particular, probabilistic model of the causal module; the causal theory, as a syntax category, will express the causal module more abstractly, to which we now turn.

3.2 Causal theories as free Markov categories

We now follow [1] to define the causal theory induced by a DAG $G = (V, A, s, t)$, a category denoted as $\mathrm{Cau}(G)$. The objects are given by words over V. A *word* over V is a finite sequence of elements of V, and this also includes the empty word $\emptyset$. Let $W(V)$ be the set of words over V. Obviously $W(V)$ is closed under *concatenation*: if $v, w \in W(V)$ then $vw \in W(V)$. So concatenation provides a monoidal product on $W(V)$, with the empty word $\emptyset$ as the unit.

Terminology 3.1. For convenience, elements of $W(V)$ are also referred to as *variables* and those of length 1, that is, the vertices in V, are called *atomic variables*. To ease the notation, we will henceforth denote all variables with lower case letters. Concatenation of two variables v, w is also written as $v \otimes w$.

An atomic variable v is a *path ancestor* of an atomic variable w if there is a directed path in G from v to w, and is an *ancestor* of w if it is a path ancestor of w or

is identical with w.

If no atomic variable occurs more than once in a variable v then v is *singular*; in particular, $\emptyset$ is singular. A singular variable is *maximal* if each atomic variable occurs exactly once in it.

Let $v = \bigotimes_{1\leq i\leq n} v_i$, where each v_i is atomic. Let $v_S = \bigotimes_{i\in S} v_i$ for $S \subseteq \{1,\dots,n\}$; set $v_\emptyset = \emptyset$. Write $w \subseteq v$, or $w \in v$ if w is atomic, and say that w is a *sub-variable* of v if w is equal to some v_S. Let $w = v_S$ and $w' = v_{S'}$. Write $v/w = v_{\{1,\dots,n\}\setminus S}$, $w \cap w' = v_{S\cap S'}$, $w \setminus w' = v_{S\setminus S'}$, and so on. We say that w, w' are *disjoint* if no atomic variable occurs in both of them. Note that being disjoint is not the same as $w \cap w' = \emptyset$, unless v is singular.

The morphisms in $\mathrm{Cau}(G)$ are generated from two distinct classes of generators (basic morphisms), in addition to the identity morphisms:

- The first class consists of duplicate and discard morphisms for each atomic variable v

$$\emptyset \xleftarrow{\epsilon_v} v \xrightarrow{\delta_v} vv,$$

  As mentioned previously, duplicate morphisms are needed to express the same variable entering multiple causal processes, and discard morphisms are needed to express ignoring or marginalizing over some outcomes of a causal process.
- The second class is the heart of the matter and consists of a *causal mechanism* for each atomic variable v

$$\mathrm{pa}(v) \xrightarrow{\kappa_v} v,$$

  where $\mathrm{pa}(v)$ is a chosen singular variable that contains all the parents of v, and is more accurately denoted by $\mathrm{pa}_G(v)$ if necessary. If $\mathrm{pa}(v) = \emptyset$ then this is just $\emptyset \longrightarrow v$, which shall be called a *exogenous* causal mechanism.

The causal theory $\mathrm{Cau}(G)$ is the *free* Markov category generated from these two classes of morphisms (and the identity morphisms), by taking all compositions and products as depicted in (2.1) and (2.3), subject only to the constraints in axioms (2.5)-(2.7).

Note that $\epsilon_\emptyset = \delta_\emptyset = 1_\emptyset$ in any Markov category. Also write $\kappa_\emptyset = 1_\emptyset$.

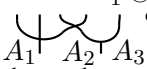
A free category is a category generated from certain generators by well-defined operations in a "no junk no noise" manner: "no junk" in the sense that only those morphisms that can be so generated are in the category, and "no noise" in the sense that no relations between morphisms hold unless they are required by the axioms. For precise technical definitions and graphical constructions of free monoidal categories, see [12]. A graphical construction of free Markov categories takes a little more work, which can be found in [15]. For the present purposes, we need not enter the rather technical details of the constructions, and we will simply use some lemmas from [15] in some of our proofs.

Remark 3.2. It may seem that the construction of $\mathrm{Cau}(G)$ depends on the choice of the singular variables $\mathrm{pa}(v)$ for the causal mechanisms κ_v. But this is not so: two distinct choices of $\mathrm{pa}(v)$ (and hence of κ_v) only differ by a permutation of atomic variables in $\mathrm{pa}(v)$ and the resulting free Markov categories are isomorphic.

Following [1] (see also [6]), we take $\mathrm{Cau}(G)$ as an categorical embodiment of the syntax for causal reasoning with G. It can then be interpreted in any Markov category via strong Markov functors, yielding different kinds of causal models. For example, a CBN based on G is a model of $\mathrm{Cau}(G)$ in the Markov category $\mathrm{FinStoch}$, the category containing stochastic matrices as morphisms ([1, 2, 6]), whereas a deterministic structural equation model based on G is a model of $\mathrm{Cau}(G)$ in the Markov category Set. We may also explore less studied causal models, such as possibilistic ones, which are models of $\mathrm{Cau}(G)$ in the Markov category Rel, the morphisms in which are relations between sets. ([1])

3.3 Causal effect morphisms

Recall the interventional probability distributions $P(\mathbf{V}|do(\mathbf{T}))$ in the context of CBN, which is often referred to as the causal effect of $\mathbf{T}$ on $\mathbf{V}$. ([14]) We now construct a class of morphisms in a causal theory that is a syntactic counterpart to such causal effects.

Notation 3.3. In any Markov category such as $\mathrm{Cau}(G)$, a morphism $A \longrightarrow B$ is called a *multiplier* on the monoidal product $A = \bigotimes_i A_i$ if it is generated from the duplicates, discards, symmetries, and identities on the factors A_i; so B must be a monoidal product of copies of the factors A_i. If $A_i \neq A_j$ for $i \neq j$ then the multiplier is unique, which is denoted by $\iota_{A\to B}$. This is due to (2.5) or, more intuitively, coherence of the graphical language for Markov categories (see [15]). For instance, if $A = A_1 \otimes A_2 \otimes A_3$ and $B = A_1^2 \otimes A_2 \otimes A_1 \otimes A_2$ then $\iota_{A\to B}$ may be depicted as 

, where how the duplicates in the trident are arranged, how the edges at the nodes are ordered, how the copies of the same object in the codomain are ordered, and so on, can all be left unspecified.

Henceforth we work in $\mathrm{Cau}(G)$.

Terminology 3.4. By the construction in [15], a morphism in $\mathrm{Cau}(G)$ is an equivalence class of string diagrams up to surgeries. Therefore, by a string diagram of a morphism, we mean any diagram in the equivalence class in question.

Notation 3.5. Although, for our purpose, there is no need to distinguish between wv and vw in $\mathrm{Cau}(G)$, for a technical reason (symmetries in free Markov categories cannot be identities), we cannot work with the quotient of $W(V)$ with respect to the relation $wv = vw$ on words. This is also the reason why the maneuver in Remark 3.2 is needed.

To remedy this, we first choose a total ordering on V and denote the corresponding maximal singular variable by $\dot{V} = 1 \otimes \ldots \otimes n$. All singular variables we shall speak of are sub-variables of $\dot{V}$. If v, w are singular variables then $v \cup w$ is abbreviated as vw or wv, which denotes the unique sub-variable of $\dot{V}$ that contains exactly the atomic variables in v, w.

The results below depend on the chosen ordering only because taking monoidal products of atomic variables does.

Definition 3.6. For singular variables t and v, let $G_{t\to v}$ be the subgraph of G that consists of all the vertices in $t \cup v$ and all the directed paths that end in v but do not *travel toward* t, that is, do not pass through or end in t (starting in t is allowed). Note that, for every $i \in V(G_{t\to v})$, if $i \notin t$ then its parents in G are all in $G_{t\to v}$ as well and if $i \in t$ then it has no parents in $G_{t\to v}$.

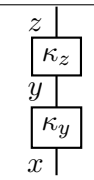
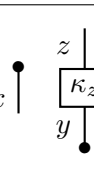
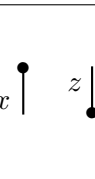
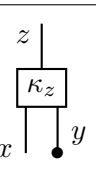
Construct a string diagram as follows. For each $i \in V(G_{t\to v})$, let $\bar{i}$ be the monoidal product of as many copies of i as the number of children of i in $G_{t\to v}$. Let Γ_i be a string diagram of

$$\begin{cases} \iota_{i\to\bar{i}\otimes i} & \text{if } i \in t \cap v, \\ \iota_{i\to\bar{i}} & \text{if } i \in t \smallsetminus v, \\ \iota_{i\to\bar{i}\otimes i} \circ \kappa_i & \text{if } i \in v \smallsetminus t, \\ \iota_{i\to\bar{i}} \circ \kappa_i & \text{if } i \notin t \cup v; \end{cases}$$

note the extra copy of i in the codomain of $\iota_{i\to\bar{i}\otimes i}$. According to Notation 3.3, there is no need to choose orderings for the codomains of the multipliers employed here. For $j \in V(G_{t\to v})$, let o_j be the number of output wires of Γ_j and p_j that of all the input wires labelled by j of all the other Γ_i. Observe that $o_j = p_j + 1$ if $j \in v$ and $o_j = p_j$ in all other cases. So we may connect the corresponding wires and fuse these *components* Γ_i into a single string diagram, denoted by $\Gamma_{[v\|t]}$, whose input wires are labelled by t and the output wires by v. The string diagram thus obtained may not be unique up to isomorphisms, but the morphism it represents is, due to coherence of the graphical language for Markov categories. This morphism is referred to as the *causal effect* of t on v and is denoted by $[v\|t] : t \longrightarrow v$, or simply $[v]$ when $t = \emptyset$, which is also called the *exogenous effect* on v.

This class of morphisms was called "causal conditionals" in [1, § 4]. We prefer to call them "causal effects" here because of their eponymous counterparts in probabilistic causal modeling mentioned earlier, but also because we will study a notion of a conditional in the next section, and $[v\|t]$ may not be a conditional in that sense.

Example 3.7. For any atomic variable v, if $\mathrm{pa}(v) = \emptyset$ then κ_v is simply depicted as  . The simplest causal effects are the ones of the form $[v\|\,\mathrm{pa}(v)]$, which is of course just the causal mechanism κ_v. Below are some simple examples of the causal effect $[z\|x]$ in $\mathrm{Cau}(G)$ for four different graphs with three vertices:

The graph G				
The subgraph $G_{x\to z}$				
The causal effect $[z\|\|x]$				

Note that in the second and third examples, x is not an ancestor of z in the causal graph G, and the causal effect $[z\|x]$ is accordingly "disconnected": the morphism factors through the monoidal unit $\emptyset$, which marks the lack of influence of x on z.

4 The Existence of Conditionals and a Generic Do-calculus

With string diagrams, we can use topological notions to aid reasoning. Here are some notions that will be used in some of the arguments below:

Definition 4.1. Let Γ be a string diagram in a symmetric monoidal category C. Denote the source of an edge e in Γ by $e(0)$ and the target by $e(1)$. A *directed path* $a \rightsquigarrow b$ in Γ is a sequence $p = (a = e_0, e_1, \dots, e_k, b = e_{k+1})$ of edges in Γ such that, for each i, $e_i(1) = e_{i+1}(0)$ and e_i is not labeled by **I**; we also regard the source and target of each e_i as belonging to the directed path, and write $p(0) = a(0)$ and $p(1) = b(1)$. A *path* is just a concatenation of finitely many directed paths. In particular, a *splitter path* (respectively, a *collider path*) is a concatenation of two directed paths joined at the starting nodes (respectively, at the ending nodes).

Two (not necessarily distinct) edges are *connected* in Γ if there is a path between them. More generally, two sets A, B of edges are connected in Γ if some $a \in A$ is connected with some $b \in B$, of particular interest is the case $A = \mathrm{dom}(\Gamma)$ and $B = \mathrm{cod}(\Gamma)$.

In some proofs below, we shall need the theory on surgeries on string diagrams developed in [15]. For Markov categories, there are four types of surgeries, corresponding to the four diagrams in (2.5) and (2.7), which shall be accordingly referred to as coassociativity surgery, counitality surgery, cocommutativity surgery, and discard surgery, respectively. Only the following bit from that theory is needed here.

Recall Remark 2.2 and Terminology 3.4. Let Γ be a string diagram of a morphism in $\mathrm{Cau}(G)$. A node x of Γ is *decorative* if either it is a dead-end (discard) or every maximal directed path p in Γ with $p(0) = x$ runs into a dead-end or, in case that x is a duplicate, this is so for all such paths through one of the prongs. Denote the set of decorative nodes of Γ by Δ_Γ and its complement by $\tilde{\Delta}_\Gamma$. Denote by P_Γ the set of the

directed paths that end in $\mathrm{cod}(\Gamma)$ and by S_Γ the set of splitter paths between edges in $\mathrm{cod}(\Gamma)$.

Lemma 4.2. *Suppose that* Γ, Υ *are string diagrams of the same morphism in* $\mathrm{Cau}(G)$. *Then*

- *There is a bijection* $\pi : \tilde{\Delta}_\Gamma \longrightarrow \tilde{\Delta}_\Upsilon$ *compatible with the labels in* Γ, Υ.
- *There is a bijection* $\dot{\pi} : P_\Gamma \longrightarrow P_\Upsilon$ *compatible with* π, *that is,* π *restricts to a bijection between the nodes in* $\tilde{\Delta}_\Gamma$ *belonging to* $p \in P_\Gamma$ *and those in* $\tilde{\Delta}_\Upsilon$ *belonging to* $\dot{\pi}(p)$.
- *There is a bijection* $\ddot{\pi} : S_\Gamma \longrightarrow S_\Upsilon$ *compatible with* π.

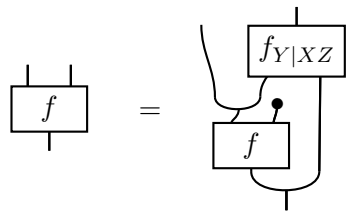
We continue to work in $\mathrm{Cau}(G)$. Here is a useful fact from [15] that will be needed in the subsequent arguments:

Lemma 4.3. *Let* $f : v \longrightarrow w$ *be a morphism between singular variables. Then there is a morphism* $g : v/v' \longrightarrow w$ *for some sub-variable* $v' \subseteq v$ *such that* $f = g \otimes \epsilon_{v'}$ *and, in all string diagrams of* g, *every atomic variable in* v/v' *is connected with an atomic variable in* w *via a directed path.*

We now proceed to establish some results about causal effect morphisms in a causal theory. A central result has to do with the existence of conditionals in a Markov category, as is defined in [2].

Definition 4.4. Let $f : Z \longrightarrow X \otimes Y$ be a morphism in a Markov category M.

- The *marginal* $f_{X|Z}$ of f over Y is the morphism  $= (1_X \otimes \epsilon_Y) \circ f : Z \longrightarrow X$.
- f admits a *conditional over* X if there is a morphism $f_{Y|XZ} : X \otimes Z \longrightarrow Y$ such that



(4.1)

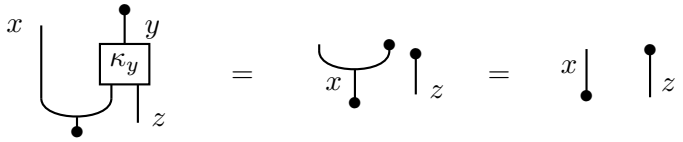
The marginals of a causal effect morphism behave as expected.

Lemma 4.5. *Let* u, v, *and* w *be singular variables with* $v \cap w = \emptyset$. *Then the marginal of the causal effect* $[vw\|u]$ *over* v *is the causal effect* $[w\|u]$ (*and that over* w *is* $[v\|u]$).

Proof. By induction on the cardinality of v, this is immediately reduced to the case where v is an atomic variable. We examine how composing with ϵ_v changes the component Γ_v and other subsequently impacted components Γ_i without changing the

morphism represented (recall Definition 3.6). If Γ_v has more than one output wire then, by counitality surgery, $\iota_{v\to\bar{v}\otimes v}$ is changed to $\iota_{v\to\bar{v}}$ and Γ_v is thus changed without impacting any other Γ_i. If Γ_v has only one output wire then, by discard surgery, it is reduced to $\bigotimes_{j\in\mathrm{pa}(v)}\epsilon_j$, where $\mathrm{pa}(v)$ is computed in $G_{u\to vw}$. In that case, for any $j\in\mathrm{pa}(v)$, we ask the same question that whether Γ_j has more than one output wire or not, and proceed accordingly as before. Observe that, when there are no more surgeries to be performed, the remaining components Γ_j, including the modified ones, are exactly those required to construct $\Gamma_{[w\|u]}$. The lemma follows. □

Note that this lemma relies on the axiom (2.7), or discard surgery, which is not imposed in [1] as we do here. To see it, consider again the third graph in Example 3.7 and the marginal $(1_x\otimes\epsilon_y)\circ[xy\|z]$ of the causal effect $[xy\|z]$ over y, then we have:



(4.2)

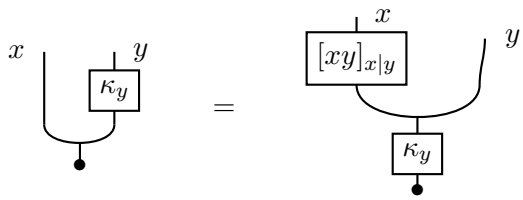
where the first equality is begotten by discard surgery and the second one by counitality surgery. Without (2.7), the first equality would fail.

Related to this observation is a claim in [1, Proposition 4.2] that if v, w are atomic variables and v is not an ancestor of w in G, then there exists no morphism $f:v\longrightarrow w$ in $\mathrm{Cau}(G)$ such that v and w are connected. Again, this is not quite right without (2.7), as shown by the example in (4.2). Now that we have imposed (2.7), this claim does hold, and is immediate from Lemma 4.3:

Proposition 4.6 ([1])**.** *Let v, w be atomic variables. If there exists a morphism $v\longrightarrow w$ in which v, w are connected then v is an ancestor of w. Conversely, if v is an ancestor of w then they are connected by a directed path in $\Gamma_{[w\|v]}$.*

This fact signals that $\mathrm{Cau}(G)$ is "purely causal", in that all connected morphisms in the category go from causal ancestors to descendants. As a result, merely "associational" or "evidential" relations are not expressed by any morphism in the category. (Recall the "no junk" property of a free category.)

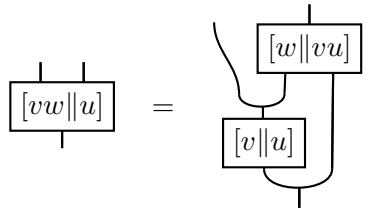
In some Markov categories such as $\mathrm{FinStoch}$ mentioned earlier, every morphism of the form $f:Z\longrightarrow X\otimes Y$ admits conditionals over both objects in the codomain ([2]), but this is not the case in $\mathrm{Cau}(G)$. Take, for instance, the simple graph $x\to y$ and consider the exogenous effect $[xy]:\emptyset\longrightarrow xy$. If a conditional $[xy]_{x|y}:y\longrightarrow x$ over y existed then we would have



(4.3)

Since the duplicate δ_y does not occur on the left-hand side, by the first claim of Lemma 4.2, its displayed occurrence on the right-hand side must be decorative, but then x, y cannot be connected by a splitter path, violating the third claim of Lemma 4.2. So the equality is not possible. On the other hand, $[xy]$ obviously admits a conditional over x, which is just $[y\|x] = \kappa_y$.

This simple example actually illustrates a general fact: for pairwise disjoint singular variables in $\mathrm{Cau}(G)$, u, v, w, if $[vw\|u]$ admits a conditional over v, the conditional must be $[w\|uv]$. We will leave the proof of this fact to another occasion, since it is a little involved and not directly relevant to the intended contributions of this paper. For the present purpose, the directly relevant question is when $[w\|uv]$ is a conditional of $[w\|uv]$, or in other words, when the following *decomposition* or *disintegration* of a causal effect holds:



(4.4)

Call the property expressed by (4.4) the *decomposability* of $[vw\|u]$ over v. We now introduce some graphical conditions relevant to characterizing decomposability and other significant concepts to be introduced presently:

Definition 4.7. Let i, j be two distinct vertices in G. A *forward trek* from i to j in G is a directed path from i to j. A *backward trek* from i to j is a directed path from j to i, or a disjoint union of two directed paths joined at a distinct starting vertex k (i.e., $i \leftarrow \cdots \leftarrow k \rightarrow \cdots \rightarrow j$).

Given $X, Y \subseteq V(G)$, a *proper* forward (respectively, backward) trek from X to Y is a forward (respectively, backward) trek from some $i \in X$ to some $j \in Y$ that does not contain any other vertex in X or in Y.

We say that

- X is *forward-t-separated* from Y by Z if every proper forward trek in G from X to Y contains some $k \in Z$;
- X is *backward-t-separated* from Y by Z if every proper backward trek in G from X to Y contains some $k \in Z$;
- X and Y are *t-separated* by Z if X is both forward-t-separated and backward-t-separated from Y by Z.

Observe that t-separation is a symmetric relation, but forward-t-separation and backward-t-separation are not. Also note that t-separation is a simpler condition than the well-known d-separation ([9]); the former is concerned only with blocking treks,

whereas the latter also has an explicit requirement for paths that contain colliders, where two arrows point at the same vertex (i.e., $i \to k \leftarrow j$).

For the rest of this section, let u, v, and w be pairwise disjoint singular variables in $\mathrm{Cau}(G)$.

Theorem 4.8. *$[vw\|u]$ is decomposable over v if and only if v is backward-t-separated from w by u.*

Proof. For the "if" direction, suppose v is backward-t-separated from w by u, and we show that the equality (4.4) holds. We examine the components Γ_i, Γ_j for $i \in V(G_{u\to v})$, $j \in V(G_{uv\to w})$ and show that, together with δ_v and δ_u on the righthand side of (4.4), they are exactly those needed to construct the string diagram $\Gamma_{[vw\|u]}$. There are the following cases.

- $i \notin uv$. Then there is a directed path from i to some $i' \in v$ in G that does not pass through u. If i also occurs in $G_{uv\to w}$ then either there is a directed path from it to some $j' \in w$ in G that does not pass through uv, or it is contained in w, both of which are prohibited by the backward-t-separation of v from w by u. So Γ_i is the same in $\Gamma_{[vw\|u]}$ and $\Gamma_{[v\|u]}$.
- $i \in uv$. Let i', j' be any children of i in $G_{u\to v}$, $G_{uv\to w}$, respectively. So $i' \notin u$ and $j' \notin uv$. If $i' \notin v$ then, by the case just considered, i' does not occur in $G_{uv\to w}$ at all; for the same reason, j' does not occur in $G_{u\to v}$. On the other hand, if $i' \in v$ then it cannot be a child of i in $G_{uv\to w}$. So Γ_i in $\Gamma_{[vw\|u]}$ is the juxtaposition of the two Γ_i in $\Gamma_{[v\|u]}$ and $\Gamma_{[w\|uv]}$ joined by δ_u.
- $j \notin uv$. Then j is an ancestor of some $j' \in w$ in $G_{(uv\to w}$, and hence cannot occur in $G_{u\to v}$, again due to the backward-t-separation of v from w by u. So Γ_j is the same in $\Gamma_{[vw\|u]}$ and $\Gamma_{[w\|uv]}$.

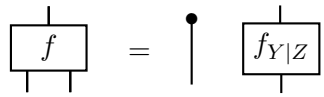
This establishes (4.4).

For the "only if" direction, suppose that (4.4) holds. Let π be a proper backward trek from $a \in v$ to $b \in w$ in G. Suppose for contradiction that π does not contain any vertex in u. By the first claim of Lemma 4.2, κ_b cannot occur in $\Gamma_{[v\|u]}$. Thus, by the other two claims of Lemma 4.2, π would translate into a directed path $b \rightsquigarrow a$ in $\Gamma_{[w\|vu]}$, which is not possible, or a splitter path between a and b in $\Gamma_{[vw\|u]}$ that does not pass through v in the direction of b and hence must pass through u, contradiction again. □

Readers familiar with Pearl's do-calculus ([10]) may have noticed the close affinity between backward-t-separation and the condition for Rule 2 of the do-calculus. Before we elaborate on the connection, let us introduce two more notions to fully match the do-calculus. One of them ("conditional independence") is introduced in [2] for all Markov categories.

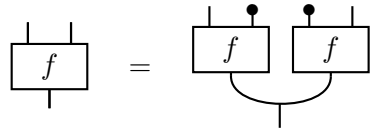
Definition 4.9. Let M be a Markov category.

- Let $f : X \otimes Z \longrightarrow Y$ be a morphism in M. We say that X is *conditionally irrelevant to Y given Z over f* if there is a morphism $f_{Y|Z} : Z \longrightarrow Y$ such that



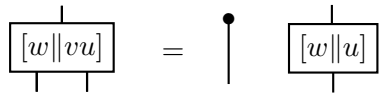
(4.5)

- Let $f : Z \longrightarrow X \otimes Y$ be a morphism in M. We say that X, Y are *conditionally independent given Z over f* if



(4.6)

We consider a specialized version of conditional irrelevance: v is *causally screened-off* from w by u if



(4.7)

It is easy to show that causal screening-off is captured precisely by forward-t-separation.

Theorem 4.10. *$[w\|vu] = \epsilon_v \otimes [w\|u]$ if and only if v is forward-t-separated from w by u in G.*

Proof. For the "if" direction, since v is forward-t-separated from w by u, no $i \in v$ can have children in $G_{vu\to w}$ and hence $G_{vu\to w}$ is the union of $G_{u\to w}$ and the trivial graph with vertices in v. It then follows from the construction of causal effects in Definition 3.6 that (4.7) holds.

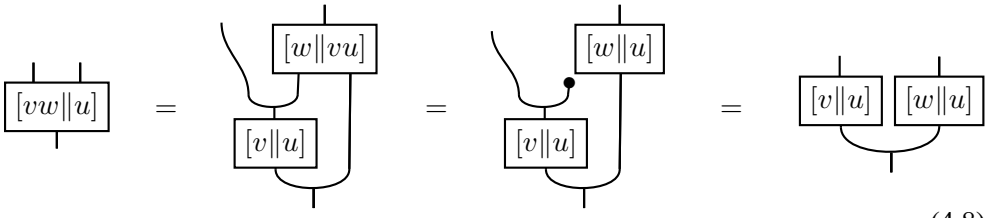
For the "only if" direction, suppose that (4.7) holds. If there is a proper forward trek from v to w in G that does not contain any vertex in u then, by Lemma 4.2, it would translate into a directed path on the right-hand side of (4.7) connecting v and w, which is not possible. □

Similarly, conditional independence over causal effects is captured precisely by t-separation.

Theorem 4.11. *We have that v, w are conditionally independent given u over $[vw\|u]$ if and only if they are t-separated by u in G.*

Proof. For the "if" direction, note that t-separation between v and w by u entails that, on the one hand, v is backward-t-separated from w by u and hence, by Theorem 4.8,

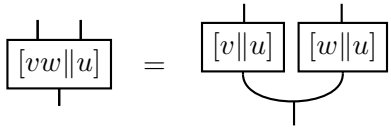
the equality (4.4) holds, and on the other hand, v is forward-t-separated from w by u and hence, by Theorem 4.10, the equality (4.7) holds. It then follows that



(4.8)

So, by Lemma 4.5, v, w are conditionally independent given u over $[vw\|u]$.

For the "only if" direction, by Lemma 4.5 again, we may assume



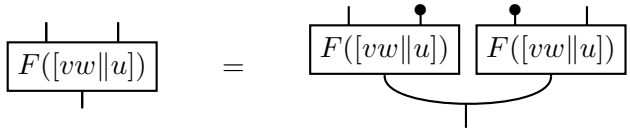
(4.9)

Let π be a proper forward or backward trek from $a \in v$ to $b \in w$. If π does not contain any vertex in u then it would translate into a directed or splitter path γ between a and b in $\Gamma_{[vw\|u]}$. Since $a, b \notin u$, we see that κ_a, κ_b occur exactly once in $\Gamma_{[vw\|u]}$ and hence, by (4.9) and the first claim of Lemma 4.2, κ_a and hence a do not occur in $\Gamma_{[w\|u]}$, whereas κ_b and hence b do not occur in $\Gamma_{[v\|u]}$. It follows from the other two claims of Lemma 4.2 that γ has to pass through u, contradiction. □

A merit of such theorems about the syntax category is that the sufficiency claims in them are immediately transferred to all models.

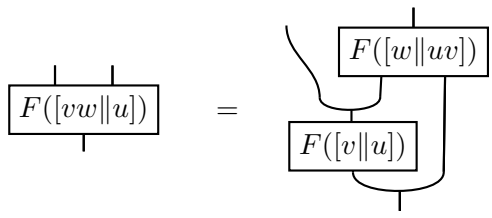
Corollary 4.12. *Let M be a Markov category and $F : \mathrm{Cau}(G) \longrightarrow \mathrm{M}$ a strong Markov functor. Then*

1. *If v and w are t-separated by u in G then, in* M*, $F(v)$ and $F(w)$ are conditionally independent given $F(u)$ over $F([vw\|u])$:*



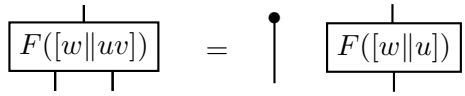
(4.10)

2. *If v is backward-t-separated from w by u in G then, in* M*, $F([w\|uv])$ is a conditional of $F([vw\|u])$:*



(4.11)

3. *If v is forward-t-separated from w by u in G then, in* M, $F(v)$ *is conditionally irrelevant to* $F(w)$ *given* $F(u)$ *over* $F([w\|uv])$:



(4.12)

On the other hand, the necessity claims do not hold in all models. For example, the decomposition property (4.11) always holds in deterministic structural equation models (as functors from $\mathrm{Cau}(G)$ to Set), regardless of backward-t-separation. The necessity in question is necessity for validity ("true in all models"), rather than truth in particular models.

5 The Causal Core of the Do-calculus

Corollary 4.12 is particularly interesting because its three clauses correspond to the three rules in Pearl's do-calculus, respectively. Suppose $\mathrm{M} = \mathrm{FinStoch}$, and F sends each causal mechanism κ_v to a positive stochastic matrix, so that it gives rise to a causal Bayesian network (CBN) model ([1]), in which the pre-intervention joint probability distribution is positive (which is assumed by Pearl's do-calculus). Let $\mathbf{X}$ denote the set of random variables represented by the object u, $\mathbf{Y}$ by w, and $\mathbf{Z}$ by v. Then equations (4.10)–(4.12) can be reformulated as follows.[2]

Equation (4.10) is rendered as:

$$P(\mathbf{Y}, \mathbf{Z}|do(\mathbf{X})) = P(\mathbf{Y}|do(\mathbf{X}))P(\mathbf{Z}|do(\mathbf{X})). \tag{5.1}$$

Equation (4.11) is rendered as:

$$P(\mathbf{Y}, \mathbf{Z}|do(\mathbf{X})) = P(\mathbf{Z}|do(\mathbf{X}))P(\mathbf{Y}|do(\mathbf{X}), do(\mathbf{Z})). \tag{5.2}$$

Equation (4.12) is rendered as:

$$P(\mathbf{Y}|do(\mathbf{X}), do(\mathbf{Z})) = P(\mathbf{Y}|do(\mathbf{X})). \tag{5.3}$$

[2]In $\mathrm{FinStoch}$, we can simply use nonzero natural numbers as the objects, and the morphisms are stochastic matrices: a morphism $n \longrightarrow m$ is a $m \times n$ stochastic matrix. A discard morphism $\epsilon_n : n \longrightarrow 1$ is the $1 \times n$ stochastic matrix in which each entry is 1, and a duplicate morphism $\delta_n : n \longrightarrow n^2$ is the $n^2 \times n$ stochastic matrix in which $(\delta_n)_i^{(i-1)n+i} = 1$ (and other entries are zero). The composition of morphisms is given by matrix multiplication, and the monoidal product is given by the Kronecker product of matrices. ([1, 2]) As shown in [1], $F([v_G])$, the image of the exogenous effect on v_G in $\mathrm{FinStoch}$, is a stochastic matrix (in fact, a column vector) encoding a joint probability distribution over the set of random variables (V) represented by v_G that satisfies the factorization in (3.1). His argument can be generalized to show that $F([v\|u])$ is a stochastic matrix encoding the distributions over the random variables represented by v given that those represented by u are intervened to take various values, according to the intervention principle (3.2). Note that equations (5.1)-(5.3) are understood as holding for all values of $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and so express equality between specific entries in the relevant matrices.

Note in addition that by the probability calculus and the assumed positivity, (5.1) is equivalent to

$$P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z}) = P(\mathbf{Y}|do(\mathbf{X})),$$

and (5.2) is equivalent to

$$P(\mathbf{Y}|do(\mathbf{X}),do(\mathbf{Z})) = P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z}),$$

because by the chain rule of the probability calculus, $P(\mathbf{Y},\mathbf{Z}|do(\mathbf{X})) = P(\mathbf{Z}|do(\mathbf{X}))$ $P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z})$.

The upshot is that Corollary 4.12, when applied to a CBN model based on G (with a positive or regular pre-intervention distribution), entails the following rules.

Rule 1 (Insertion/deletion of observations): if $\mathbf{Y}$ and $\mathbf{Z}$ are t-separated by $\mathbf{X}$ in G, then

$$P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z}) = P(\mathbf{Y}|do(\mathbf{X})).$$

Rule 2 (Action/observation exchange): if $\mathbf{Z}$ is backward-t-separated from $\mathbf{Y}$ by $\mathbf{X}$ in G, then

$$P(\mathbf{Y}|do(\mathbf{X}),do(\mathbf{Z})) = P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z}).$$

Rule 3 (Insertion/deletion of actions): if $\mathbf{Z}$ is forward-t-separated from $\mathbf{Y}$ by $\mathbf{X}$ in G, then

$$P(\mathbf{Y}|do(\mathbf{X}),do(\mathbf{Z})) = P(\mathbf{Y}|do(\mathbf{X})).$$

Pearl's do-calculus ([10]) consists of exactly three rules like these, but each rule therein is more general than the corresponding rule above and is formulated in terms of the more complicated d-separation criterion and various modifications of G. The extra generality in Pearl's version is that the consequent equation in each rule has an extra set of variables $\mathbf{W}$ to be conditioned upon on both sides of the equation. For example, the consequent equation in the first rule of Pearl's do-calculus is

$$P(\mathbf{Y}|do(\mathbf{X}),\mathbf{Z},\mathbf{W}) = P(\mathbf{Y}|do(\mathbf{X}),\mathbf{W}),$$

and similarly for the other two rules. It is a simple exercise to check that each of the rules above is exactly equivalent to the corresponding rule in Pearl's calculus when $\mathbf{W}$ is taken to be empty.

So Corollary 4.12, when applied to a CBN model, yields a specialized version of Pearl's do-calculus. However, although each rule in the specialized version is a special case of the corresponding rule in the full version, taken together they are actually as strong as the full version. To see this, it suffices to show that we can recover the intervention principle (3.2) from the specialized rules, or to be more exact, from Rule 2 and Rule 3 above, since Rule 1 is entailed by the conjunction of Rule 2 and Rule 3 (just as in the full version, see [5]). Since the full version is entailed by the intervention principle (plus the probability calculus), it is also entailed by the specialized

version (plus the probability calculus) if the intervention principle is entailed by the specialized version (plus the probability calculus).

We now sketch a fairly simple argument to that effect. It helps to first consider the pre-intervention case, where we need to show that Rule 2 and Rule 3 entail that the pre-intervention probability distribution factorizes as in (3.1). This is equivalent to deriving the local Markov condition from Rule 2 and Rule 3, the condition that every variable is probabilistically independent of its non-descendants conditional on its parents. ([13]) The derivation is straightforward. For any random variable $V \in \mathbf{V}$ in the CBN, by Rule 2, we have

$$P(V|do(\text{pa}(V))) = P(V|\,\text{pa}(V)),$$

and

$$P(V|do(\text{nd}(V))) = P(V|\,\text{nd}(V)),$$

where $\text{pa}(V)$ and $\text{nd}(V)$ denote the set of V's parents in G and the set of V's non-descendants in G (i.e., those variables of which V is not an ancestor), respectively. This is so because $\text{pa}(V)$ is trivially backward-t-separated from V by the empty set (for there is no proper backward trek from $\text{pa}(V)$ to V), and so is $\text{nd}(V)$, which contains $\text{pa}(V)$ as a subset. Then by Rule 3, we have

$$P(V|do(\text{nd}(V))) = P(V|do(\text{pa}(V))),$$

simply because every forward trek to V contains a parent of V. It follows that for each V,

$$P(V|\,\text{nd}(V)) = P(V|\,\text{pa}(V)).$$

So the factorization required by the intervention principle holds in the pre-intervention case. This argument easily generalizes to any post-intervention probability distribution $P(\mathbf{V}|do(\mathbf{T}))$; that is, we can derive in the same fashion from Rule 2 and Rule 3 that for every V,

$$P(V|\,\text{nd}^*(V), do(\mathbf{T})) = P(V|\,\text{pa}^*(V), do(\mathbf{T})),$$

where $\text{pa}^*(V)$ and $\text{nd}^*(V)$ denote the set of V's parents and the set of V's non-descendants, respectively, in the subgraph of G in which all arrows into variables in $\mathbf{T}$ are deleted. From this follows the factorization of $P(\mathbf{V}|do(\mathbf{T}))$ required by the intervention principle (3.2).

Therefore, the full do-calculus can in principle be derived from the specialized do-calculus together with the probability calculus. This fact reveals an equivalent formulation of the do-calculus for CBN models that is simpler than the standard formulation. More importantly, this simpler formulation reflects the "causal core" of the do-calculus, for it is an instance of the generic do-calculus given in Corollary 4.12,

and the generic do-calculus is derived from results in a syntax category that is, so to speak, purely causal (because there is no morphism in that category to match non-causal or evidential relations between variables. More precisely, the causal core of the do-calculus is given by the rule for causal decomposition ((4.11), rendered as (5.2) in a CBN model) and the rule for causal screening-off ((4.12), rendered as (5.3) in a CBN model). These are derived without any consideration of the non-causal features of the models. The standard do-calculus for the CBN models can be seen as derived from a conjunction of these two rules on the one hand, which are purely causal, and the probability calculus on the other hand, which is non-causal.

6 Conclusion

Following the pioneering work of [1], we studied the causal effect morphisms in a causal DAG-induced free Markov category in some depth, and established sufficient and necessary graphical conditions for some conceptually important properties of such morphisms, including especially decomposition and screening-off. Our results yield a generic do-calculus that is more general and abstract than the standard do-calculus in the context of causal Bayesian networks. Not only is the generic do-calculus more widely applicable, it is also conceptually illuminating in that it reveals the purely causal component of the do-calculus. When applied to causal Bayesian networks, it also results in a simpler but equivalent formulation of the probabilistic do-calculus.

Since the simpler do-calculus uses trek-separation rather than the more convoluted d-separation, it is probably easier to explain and understand. Moreover, the simpler do-calculus may also be readily extendable to other causal graphical models derived from DAG models. For example, in [16], Pearl's do-calculus is extended to the so-called partial ancestral graphs (PAGs), which are used to represent Markov equivalence classes of DAG models. The extension is intended to capture the applicability of a do-calculus rule in all DAGs in the equivalence class represented by a PAG, but due to the complex graphical conditions in Pearl's do-calculus, it only accommodates some but not all such cases of unanimous applicability. We suspect that an extension of the simpler do-calculus highlighted in this paper would be more straightforward and complete.

References

[1] B. Fong, 2013, "Causal Theories: A Categorical Perspective on Bayesian Networks", arXiv:1301.6201.

[2] T. Fritz, 2020, "A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics", *Advances in Mathematics*, **370(3)**: 107239.

[3] D. Galles and J. Pearl, 1998, "An axiomatic characterization of causal counterfactuals", *Foundations of Science*, **3(1)**: 151–182.

[4] J. Y. Halpern, 2000, "Axiomatizing causal reasoning", *Journal of Artificial Intelligence Research*, **12(1)**: 317–337.

[5] Y. Huang and M. Valtorta, 2006, "Pearl's calculus of intervention is complete", *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence*, pp. 217–224, arXiv:1208.5160.

[6] B. Jacobs, A. Kissinger and F. Zanasi, 2019, "Causal inference by string diagram surgery", *International Conference on Foundations of Software Science and Computation Structures*, pp. 313–329.

[7] F. W. Lawvere, 1963, "Functorial semantics of algebraic theories", *Proceedings of the National Academy of Sciences of the United States of America*, **50(5)**: 869–872.

[8] S. Mac Lane, 1998, *Categories for the Working Mathematician* (2nd edition), Berlin: Springer-Verlag.

[9] J. Pearl, 1988, *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.

[10] J. Pearl, 1995, "Causal diagrams for empirical research", *Biometrika*, **82(4)**: 669–710.

[11] J. Pearl, 2009, *Causality: Models, Reasoning, and Inference* (2nd edition), Cambridge: Cambridge University Press.

[12] P. Selinger, 2010, "A survey of graphical languages for monoidal categories", *New Structures for Physics*, pp. 289–355, Berlin Heidelberg: Springer.

[13] P. Spirtes, C. Glymour and R. Scheines, 2000, *Causation, Prediction, and Search* (2nd edition), Cambridge, MA: MIT Press.

[14] J. Tian and J. Pearl, 2002, "A general identification condition for causal effects", *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pp. 567–573, Menlo Park, California: The AAAI Press.

[15] Y. Yin and J. Zhang, "A formal account of causal modularity", manuscript.

[16] J. Zhang, 2008, "Causal reasoning with ancestral graphs", *Journal of Machine Learning Research*, **9(3)**: 1437–1474.

[17] J. Zhang and P. Spirtes, 2011, "Intervention, determinism, and the causal minimality condition", *Synthese*, **182(3)**: 335–347.

马尔可夫范畴、因果理论与干预演算

尹一木　　张寄冀

摘　要

本文循范畴论的方式处理基于有向无环图的因果模型，用一种自由马尔可夫范畴（名曰“因果理论”）作为因果推理的语形范畴，并在此框架下定义因果推理中的一些重要概念，比如因果独立性、因果条件化，以及干预作用的分解。本文的结果不依赖于常见的一些因果模型的特性，因此适用的范围更广，概念上也更加清晰。它们与朱迪亚·珀尔的著名的干预演算关系密切，相当于一个语形层面的干预演算，在各种基于有向无环图的因果模型上都有效。对因果贝叶斯网这种概率因果模型来说，这个语形层面的干预演算对应于珀尔演算的一个相对简单的特殊形式。此特殊形式实际上是干预演算的“因果核心”，加上概率演算就足以推导出标准的珀尔演算。

尹一木　加利福尼亚帕萨迪纳
yimu.yin@hotmail.com
张寄冀　香港浸会大学宗教及哲学系
zhangjiji@hkbu.edu.hk