The process theory of causality: an overview

Jun Otsuka, Hayato Saigo

Abstract

This article offers an informal overview of the category-theoretical approach to causal modeling introduced by Jacobs et al. (2019) and explores some of its conceptual as well as methodological implications. The categorical formalism emphasizes the aspect of causality as a process, and represents a causal system as a network of connected mechanisms. We show that this alternative perspective sheds new light on the long-standing issue regarding the validity of the Markov condition, and also provides a formal mapping between micro-level causal models and abstracted macro models.

1 Introduction

Graphical modeling is now the standard toolkit for studying causality and finding causal relationships from observed data (Spirtes et al., 1993; Pearl, 2000). A typical causal model $\mathcal{M} = (G, P)$ in this approach consists of a directed acyclic graph (DAG) G over a set of variables and a probability distribution P, where the graph $G = (\boldsymbol{V}, \boldsymbol{E})$ is a pair of a set $\boldsymbol{V}$ of variable and a set $\boldsymbol{E} \subset \boldsymbol{V} \times \boldsymbol{V}$ of edges between them. Variables designate properties or states of units or objects, say diet or blood pressure of patients. The existence of edge from one variable to another means that a state of the latter is causally dependent on that of the former, in such a way that an intervention in the former results in a change in the latter. Thus, causality in this framework is understood as *relationship between events*, where events are designated by variables assuming particular values. For instance, *BloodPresure* = *high* would designate the event that a given patient's blood pressure is high, and a causal question asks whether this type of event is in a systematic relationships with other types of events regarding, say, one's diet or other medical conditions.

The event-centered view dates back to British Empiricism and especially to David Hume, who took inductive reasoning to be an inference from one type of events to another. For Hume, this task was equivalent to establishing a causal relationship between them, which he thought can never be warranted by logic or experience. Contemporary statistics and machine learning research has tried to alleviate this skepticism by introducing various empirical and theoretical assumptions that would allow algorithmic identification of causal relationships from observed data (Morgan and Winship, 2007; Peters et al., 2017), but the basic conceptual framework remains the same: a causal system is considered as a constellation of events/variables manifesting regular patterns.

On the other hand, some philosophers have proposed an alternative conception of causality, featuring its aspect as *processes* (Salmon, 1984; Dowe, 2000) or *mechanisms* (Machamer et al., 2000; Cartwright, 2007). Causality, according to this view, is best understood as a process that transmits influence from one event to another, or a mechanism that produces an outcome by taking some inputs. For instance, one may take a metabolic process as a mechanism that "generates" blood pressure (among other things) in response to, say, a dietary practice.

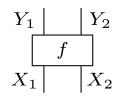
We believe that this process-centered view of causality can be formally represented using the category-theoretic language of *string diagrams*, and that this alternative formalism sheds new light on some problems of causality regarding the causal Markov condition and the problem of abstraction. Following the seminal work of Jacobs et al. (2019), section 2 presents the categorical formalization of discrete causal models with finite variables. We show how causal DAGs are translated to string diagrams, and that a functorial mapping of diagrams yields causal models. Our presentation prioritizes clarity over theoretical rigor and proceeds with examples rather than mathematical proofs so that the core idea can be grasped by a reader without familiarity with category theory. Section 3 then looks at the old problem of the Markov condition from the categorical perspective, and points out that the validity of this condition hinges on the existence of a special mechanism called *copier*, which duplicates a causal process without disturbing it. Section 4 turns to the problem of abstracting a causal model by coarsening its variables. The challenge of abstraction is to map a "low-level" micro model to a "high-level" macro model in a consistent fashion. We will show that this mapping is given by a category-theoretic notion of *natural transformation* between two causal models/functors. We conclude that the category-theoretic approach offers a novel perspective and solutions to some issues that resisted successful formal treatment in the conventional DAG formalism.

2 Process theory of causality

While the event-centered view of causality has a natural representation in graphical modeling, the process-centered view can be formalized by using *process theory*, which have been developed mainly in categorical quantum mechanics and computer science (e.g. Abramsky and Coecke, 2004; Coecke and Kissinger, 2017). Here we briefly review the application of the process theory to causal modeling introduced by Jacobs et al. (2019).

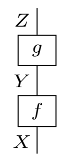
2.1 Translating a DAG to a string diagram

Process theory conceptualizes a process as a system of combined mechanisms that exchange their products with each other. Each mechanism, commonly represented by a box, has definite types of inputs and outputs, represented by wires (there are also special types of mechanisms that do not have either or both of input and output). The following is an example of a mechanism that takes two inputs X_1, X_2 and returns two outputs Y_1, Y_2;



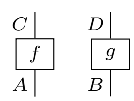
Unless mentioned otherwise, it is assumed that causal processes flow from bottom to top.

Given two boxes f and g, if the type of an output of f matches with that of an input of g, these two boxes can be combined vertically via the matching wire, as:



Intuitively this can be understood as an initial input X processed by f being transmitted for a further processing by g to yield a final outcome Z.

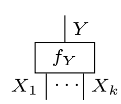
In addition to the vertical composition, multiple streams can be combined horizontally, representing a parallel processing:



This describes the situation where two types of inputs, A and B, are independently processed by f and g respectively, to output C and D. Parallel processing can also be understood as a combined input $A \otimes B$ processed by a combined process $f \otimes g$ to yield $C \otimes D$.

A system created by combining multiple mechanisms via vertical and parallel compositions is called *string diagram.* A string diagram as a whole can be considered as one big process that has combined inputs and outputs.

In the context of causal modeling, a diagram serves as a causal graph, describing the topological feature (i.e., connectedness) of a causal system. Wires in a string diagram corresponds to variables. For each variable $Y \in \boldsymbol{V}$, there is a box of the form



(1)

where $X_1, \ldots, X_k \in \mathrm{PA}(Y)$ are parents of Y. Intuitively, the box represents a "generating mechanism" of Y that takes $\mathrm{PA}(Y)$ as input, and thus multiple edges pointing to one variable are summarized by one box. In addition, we assume that an exogenous variable (with $\mathrm{PA}(Y) = \emptyset$) also has its own "state" with no input, depicted with a triangle. By combining these boxes and wires in accordance with a given DAG, one can create a matching string diagram, as illustrated in Fig. 1. Note that a string diagram gives a somewhat "flipped" image of the graph, replacing nodes with wires and edges with boxes.

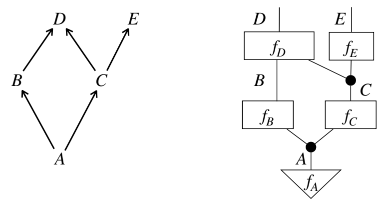
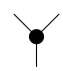


Figure 1: A translation of a DAG (left) to a string diagram (right).

One component of the string diagram in Fig 1 that lacks an explicit graph counterpart is the cloning process or *copier*:



which duplicates the input and returns two (or more) outputs of the same type. A copier is needed when there is a fork $X \leftarrow Y \rightarrow Z$ in the graph. From the process-perspective, this means that the product Y is used twice, one for an input to (the generating mechanism of) X and the other to Z. Such an operation is taken for granted in causal graphs, but not in a string diagram and must be explicitly counted as an independent process. This is because duplication is not always possible: in quantum mechanics, for instance, one can not copy one state without disturbing it. In section 3, we will discuss that the existence of copier also proves to be crucial for the validity of the causal Markov condition.

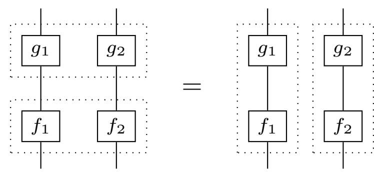
String diagrams can be formally described in the language of *symmetric monoidal category*. Wires and boxes of a string diagram are objects and morphisms (arrows) of this category. A vertical composition of boxes correspond to the composition of morphisms with the matching codomain/domain: for example the composition of $f : A \rightarrow B$ and $g : B \rightarrow C$ yields $g \circ f : A \rightarrow C$. A parallel composition is given by binary associative operations of objects and morphisms:

$$\begin{aligned}\otimes :&\mathrm{ob}(\mathcal{C}) \times \mathrm{ob}(\mathcal{C}) \rightarrow \mathrm{ob}(\mathcal{C}) \\ \otimes :&\mathcal{C}(A, B) \times \mathcal{C}(C, D) \rightarrow \mathcal{C}(A \otimes B, C \otimes D)\end{aligned}$$

where $\mathrm{ob}(\mathcal{C})$ is a class of objects and $\mathcal{C}(A, B)$ is the set ("homset") of morphisms from A to B of category $\mathcal{C}$. The vertical and parallel compositions of morphisms f_1, f_2, g_1, g_2 must be commutative:

$$(g_1 \otimes g_2) \circ (f_1 \otimes f_2) = (g_1 \circ f_1) \otimes (g_2 \circ f_2).$$

In the diagram presentation, this just means the two ways of composing processes



yield the same diagram.

On this categorical background, Jacobs et al. (2019) introduce a free category (or what they call *free CDU category*) over a pair of generating sets of objects and morphisms. In particular, a causal string category Syn_G is built from a DAG $G = (\boldsymbol{V}, \boldsymbol{E})$ by taking its variable set $\boldsymbol{V}$ as the generating set of objects and the set of boxes of the form (1) as the generating set of morphisms. That this is a free category means that Syn_G contains anything that can be obtained by combining these wires and boxes (plus some other special units such as copiers, discards, units). This includes not just the string diagram in Fig 1, but also any of its parts and their suitable combinations.

2.2 Probabilistic interpretation of a string diagram

A causal model is a probabilistic interpretation of string diagrams in the above defined free CDU category. This is done by a *functor*, a systematic mapping from one category to another, in the present case from Syn_G to a Markov category of an appropriate structure (Jacobs et al., 2019; Fritz, 2020). For discrete causal models whose variables have only finite values, the target Markov category will be $\mathsf{FinStoch}$, whose objects are finite sets and morphisms $f : X \to Y$ are $|Y| \times |X|$ dimensional stochastic matrices, i.e., matrices of positive numbers whose columns each sum up to 1. A functor then assigns each wire of Syn_G with a finite set (representing values of the corresponding variable) and each box with a stochastic matrix (representing conditional probabilities of the effect given its causes, also known as *Markov kernels*). In addition, a state (a triangle with no input) of an exogenous wire/variable X is mapped to a morphism from the object 1 of $\mathsf{FinStoch}$; this morphism is a $|X| \times 1$ stochastic matrix or vector, and thus gives a marginal distribution $P(X)$ of X.

Fig. 2 illustrates probability assignments by causal model functor F to the bottom half of the string diagram in Fig. 1. Here, each variable/wire is assumed to have two values, and thus mapped to two-element sets $\{a_1, a_2\}$, $\{b_1, b_2\}$, and $\{c_1, c_2\}$. The left-most box $F(f_A)$ gives a marginal distribution $P(A)$ in the 2×1 vector format. $F(cp_A)$ interprets the copier with a $(2 \times 2) \times 2$ matrix that effectively "duplicates" $P(A)$ to yield $P(A \times A)$. This is in turn fed into $F(f_B)$ and $F(f_C)$, 2×2 matrices each representing the conditional distribution $P(B|A)$ and $P(C|A)$, respectively. As a whole, the functor gives the joint probability distribution $P(A, B, C)$ that satisfies the Markov condition with the DAG $B \leftarrow A \to C$.[1]

[1]Precisely speaking, in string diagrams only those wires extending to the end are assumed

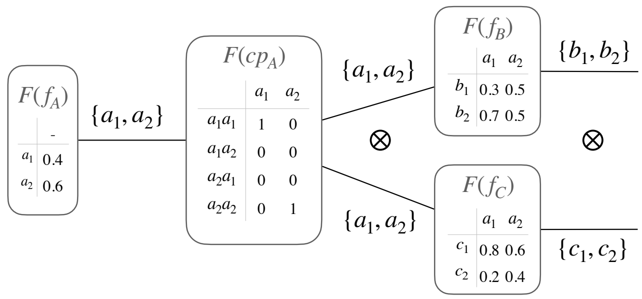
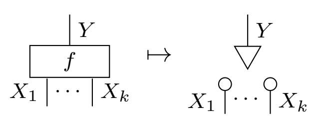


Figure 2: An example of a functorial assignment of values and (conditional) probabilities to a string diagram. Here the causal flow is from left to right. The structure shown interprets the bottom part of the string diagram in Fig. 1.

A different functor $F' : \mathsf{Syn}_G \to \mathsf{FinStoch}$ leads to different probability assignments, possibly with different numbers of values of the variables. In this way any causal model that satisfies the Markov condition with DAG G can be represented as a functor. In fact, this correspondence is one-to-one, which means that a discrete acyclic causal model (G, P) can be identified with a functor $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$ (Jacobs et al., 2019).

2.3 Intervention via diagram surgery

One of the core features of causal modeling is the operation of intervention, which forces a target variable to assume a particular distribution. In the categorical formalization, an intervention is defined as a diagram surgery which replaces any appearance of the box of a target variable with an exogenous "state" (triangle) and discards its inputs (denoted by empty circles):



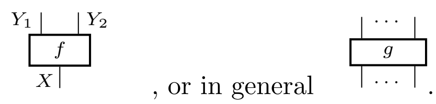
while keeping all the other boxes and wires intact. For a string diagram category Syn_G, this mapping defines an endofunctor $\mathsf{cut}_Y : \mathsf{Syn}_G \to \mathsf{Syn}_G$. Interventions on other variables give rise to similar endofunctors. A post-intervention distribution is obtained by combining an intervention functor with a causal model functor, such that $F \cdot \mathsf{cut}_Y : \mathsf{Syn}_G \to \mathsf{FinStoch}$.

to be observed. Hence to have a joint distribution $P(A, B, C)$, one needs to branch A once more and run it to the end. But in this article we ignore this convention and assume that all wires in a string diagram are observed.

3 The Markov condition

We have reviewed so far the categorical formalization of causal models by Jacobs et al. (2019) as a formal representation of the process-oriented view of causality. The advantage of taking this alternative perspective is that it sheds light on some issues that resist a proper theoretical handling in the conventional DAG formalism. Jacobs et al. (2019) showed that the identifiability of intervention outcomes can be easily determined by the diagrammatic operation called *comb disintegration*. In this and next section we discuss two other issues, one concerning the Markov condition and the other abstraction of causal models.

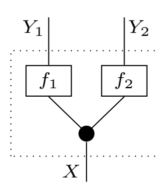
At the end of the previous section, we noted the one-to-one correspondence between a discrete causal model (G, E) and a functor $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$. This, however, does not mean the equivalence of the diagrammatic and graph-theoretic formalization. In fact, the former can deal with a broader range of causal structures. The aforementioned procedure of constructing a string diagram from a causal graph was based on the assumption that each variable/wire has its own generating mechanism, giving rise to a box with just one output. But this needs not be the case in the process theory (or symmetric monoidal category) in general: boxes may have multiple outputs, such as:



(2)

Since such boxes do not arise in the construction of Syn_G from a DAG G, they suggest the possibility of causal structures that do not have a graph-theoretical counterpart (Jacobs, 2022).

It should be emphasized that the left box f in (2) is *not* equivalent to a fork $Y_1 \leftarrow X \rightarrow Y_2$. For if it were a fork, the Markov condition should entail $Y_1 \perp Y_2 | X$. Nothing in the diagrammatic representation, however, enforces this independence relationship. The morphism f in (2) can be mapped by a functor to any stochastic matrix $P(Y_1, Y_2|X)$, where Y_1 and Y_2 may or may not be independent given X. To obtain the independence of two outputs Y_1 and Y_2, the branching must be made via a copier:

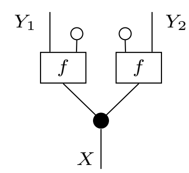


(3)

This is the correct diagrammatic rendition of the fork $Y_1 \leftarrow X \rightarrow Y_2$ in a causal DAG, which makes Y_1 and Y_2 independent given X in any functorial (probabilistic) interpretation of this diagram. Note also that since every box in this diagram has just one output, it can be constructed from a graph following Jacobs et al's procedure.

From the other perspective, the causal Markov condition can be understood as the claim that every multi-output process as in (2) is a disguised dashed box as in (3) and must be decomposable into separate mechanisms with a copier. Note that (3) implies that each of Y_1 and Y_2 can be modified without affecting the other by a diagrammatic surgery of box f_1 or f_2, whereas such a modular intervention is barred in (2). Hence, the assumption that any multi-output box as in (2) is replaceable by (3) can be properly called *modularity condition.* What was shown by the diagrammatic reasoning above is that the modularity condition in this sense does imply the Markov condition.[2]

The question, then, boils down to the validity of the modularity condition, and it is this point that critics have put under critical scrutiny (Cartwright, 1999, 2007). Cartwright argues that the Markov condition fails when a cause operates probabilistically, and illustrates her claim with a hypothetical chemical factory which generates products Y_1 and side-effect pollutants Y_2, with certain probabilities such that Y_1 and Y_2 do not become independent even conditioned on the operation of the factory (Cartwright, 2007, p. 107). Her factory is nothing but the process f in (2), and her claim is that it is not decomposable as in (3) because the chemical products and pollutants are generated by the same generating mechanism by assumption. Her argument can be paraphrased using diagrams: it can be shown that, if f in (2) is equivalent to (3), it can also be rewritten as:



(4)

where empty circles are operators that "discard" each of the two outputs Y_1, Y_2 (Fritz, 2020, Lemma 12.11). This means that modularity (3) assumes that the two outputs Y_1 and Y_2 are produced by applying the same production process f to the input X twice and then discarding one of the outputs in each. This strikes as a rather strong assumption, which is unlikely to hold in situations like Cartwright's example.

Both in (3) and (4), the copier plays a crucial role. In this sense, the crux of the Markov condition and the modularity condition is the existence of a copier: is it always possible to duplicate one process without disturbing it? The answer is known to be negative in the quantum context. The possibility in marco, non-quantum setups seems to depend, but the process view suggests an interesting empirical hypothesis. The hypothesis is that if an alleged common cause is a repeatable event or condition that generates effects at separate moments or through different mechanisms without altering its nature as a cause, as when a

[2] Hausman and Woodward (1999) proposed a similar argument for the Markov condition, but their definition of modularity just requires the invariance of structural equations to changes in values of variables, and is weaker than the modularity defined here as a diagrammatic surgery. Cartwright (2007) shows that Hausman and Woodward's modularity is insufficient to ensure the Markov condition.

genotype of an individual (e.g. possession of abnormal copies of the β-globin gene) is said to be a cause of two distinct symptoms (sickle cell disease and resistance to malaria), the modularity and the Markov condition are expected to be satisfied. If, on the other hand, the causal factor is transitory and altered, consumed, or destroyed each time it produces its effects, as in chemical reactions, these conditions are likely to fail. Confirmation or disconfirmation of this hypothesis, however, must await further empirical investigation.

4 Abstracting Causal Models

The next problem that we focus on is the problem of abstracting causal models. Causal systems can be described at different levels of granularity, and finding an appropriate macro-level causal features out of micro-level measurements (such as gene expression data or image pixels) is a major challenge in machine learning and scientific inquiries in general (Iwasaki and Simon, 1994; Chalupka et al., 2014, 2016; Schölkopf et al., 2021). The assumption of coarsening is that the models at different levels, despite having different set of variables and edges, are consistently related so that they are regarded as modeling the same phenomenon. Recent studies have proposed formal conditions of such an abstraction procedure that maps components of a finer-grained "low-level" model to those of a coarser-grained "high-level" model (Rubenstein et al. 2017; Beckers and Halpern 2019; Beckers et al. 2020; Rischel 2020; Rischel and Weichwald 2021; Otsuka and Saigo 2022; see Zennaro 2022 for review).

Coarsening may operate on variables, by merging multiple micro variables into one macro variable, or on values, by reducing multiple values of one variable to a fewer number of values with less resolution, or both (Zennaro, 2022). Either way, for the resulting model to count as an abstraction of the original model, such a mapping must be consistent in three essential aspects of causal models:

1. Structural: causal relationships of the low-level model must be preserved. In particular, if there is an edge between two micro variables, their macro counterparts must also have an edge in the matching direction.

2. Probabilistic: the probability assignment of the high-level model must be consistent with that of the low-level model.

3. Interventional: the two models must make consistent predictions to external interventions.

Another way of spelling out these desiderata is that abstraction procedure must commute with various operations in/on a causal model. For instance, the probabilistic consistency would require that the probability of an effect calculated in the micro model must "match" with that of its macro counterpart (Rubenstein et al., 2017; Rischel, 2020; Rischel and Weichwald, 2021). We illustrate below that the category-theoretic formulation provides a natural micro-macro translation that fulfills all these desiderata.

4.1 Abstraction in a monoidal category

Let us begin with the value reduction. There are two types of value reduction: the first is a deterministic transformation or *supervenience*, that merges multiple values of one variable into a fewer number of values with less resolution. For discrete variables such a map is given by a rank-deficient stochastic matrix whose entries are 1 or 0. The second type is stochastic, and simply maps one variable to another with any stochastic matrix whose size matches the number of values of the source and target variables. The categorical approach handles both types in the same way using the notion of natural transformation.

Suppose we are given a causal model $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$. An abstracted model that merges some values of its variables is given by another functor $F' : \mathsf{Syn}_G \to \mathsf{FinStoch}$ such that $|F'(X)| \leq |F(X)|$ for any object X of Syn_G. Thus an abstraction is a mapping between functors $F \Rightarrow F'$ that fulfills the consistency requirements listed above. In category theory, such a mapping is called *natural transformation*. Given two causal model functors $F, F' : \mathsf{Syn}_G \to \mathsf{FinStoch}$, a natural transformation $\alpha : F \Rightarrow F'$ is a set of morphisms in $\mathsf{FinStoch}$ (therefore stochastic matrices) that make the following diagram commute for any morphism $f : X \to Y$ in Syn_G:

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ {\scriptstyle \alpha_X}\big\downarrow & & \big\downarrow{\scriptstyle \alpha_Y} \\ F'(X) & \xrightarrow{F'(f)} & F'(Y) \end{array} \tag{5}$$

Here the upper half represents a stochastic transition along the causal edge $f : X \to Y$ according to the original model F, while the bottom represents the corresponding transition in the coarse-grained model F'. They are stochastic matrices of dimension $|F(Y)| \times |F(X)|$ and $|F'(Y)| \times |F'(X)|$, respectively. In contrast, vertical arrows α_X and α_Y serve to relate these causal flows in a consistent fashion. They are also stochastic matrices, and in case of deterministic translation (i.e., merging of values) their entries are either 1 or 0. The commutativity of the diagram means that coarsening $\alpha_X : F(X) \to F'(X)$ is consistent at every step of the causal flow, in the sense that one obtains the same marginal distribution regardless of whether one follows the causal path in the original model and then transforms the effect (clockwise path) or transforms the cause first and then derives its effect in the coarse-grained model (counterclockwise path). The existence of a natural transformation between two models/functors F and F' thus warrants the probabilistic consistency.

One difficulty of the value reduction is that it may infringe on the Markov condition. In general, there is no guarantee that conditioning on a coarser-grained redescription X' of the common cause in $Y \leftarrow X \to Z$ would make its effects independent, i.e., $Y \not\perp Z | X'$ even if $Y \perp Z | X$. The categorical approach, however, is free from this problem, because the fact that F' is a functor (i.e., causal model) ensures that it satisfies the Markov condition with the original graph.

In general, finding an abstraction between two candidate models is a non-trivial task. In the case of deterministic abstraction, however, there is a necessary and sufficient condition for the existence of a transformation (Otsuka and Saigo, 2022). This condition is called *causal homogeneity*, and intuitively requires that micro values to be merged into the same macro value must have homogeneous causal effects *modulo* groups of the effect variable. For more details, see Otsuka and Saigo (2022). Additionally, Rischel (2020); Rischel and Weichwald (2021) propose the use of KL-divergence to measure the non-commutativity of abstraction when the exact match between two models is hard to come by, as is expected in empirical measurements.

Let's now move on to the next problem of variable reduction, where two or more variables in a macro model are merged into one variable in a micro model. In a way, this type of merging is already built in a monoidal category as vertical or horizontal compositions in a string diagrams. Recall that Syn_G, as a free symmetric monoidal category, contains any appropriate compositions of the generating objects and morphisms. One may then consider such combined objects or morphisms as "abstractions" of its components. For instance, Fig. 3 shows progressive procedures by which components are combined to form larger processes, which can then be considered as abstraction of its constituting processes. The horizontal and vertical compositions of string diagram, therefore, provide a means for variable reduction. The functorial property of a causal model then takes care of the probabilistic as well as interventional consistency: in particular, the probabilistic interpretation of the merged processes can be calculated from that of its constituent.

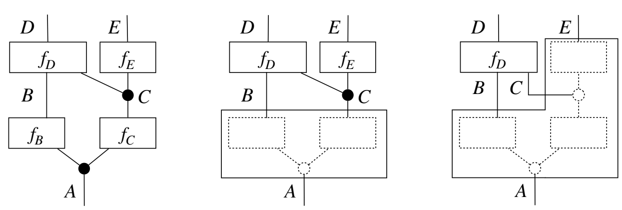
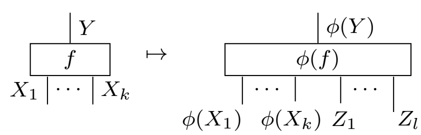


Figure 3: "Abstraction" with string diagrams. In symmetric monoidal categories, objects (wires) and morphisms (boxes) can be combined to make another objects and morphism, which can then be considered as a joint mechanism. The string diagram in the middle combines a copier and two parallel processes, f_B and f_C into one process. The inverse L-shaped box on the right further encompasses another copier and f_E, making up a process with three outputs B, C and E.

The categorical/monoidal compositions by themselves, however, cannot be considered a full-brown abstraction. Abstraction is expected not only to consolidate information, but also to discard or forget some of it. Composition may serve for the former but not the latter purpose, for composed boxes or wires retain all the details as its components. Moreover, it does not serve if our aim is to compare two causal *graphs*. Boxes resulting from compositions may have multiple outputs, in which case there is no obvious graph-theoretic counterpart (Sec. 3). For example, there is no causal graph that corresponds to the middle and right string diagrams in Fig. 3 that preserves the cause-effect relationships in the original causal graph (Fig. 1). If one wishes to see abstracted models in the conventional graphical formalism, a different approach must be taken.

4.2 Abstraction via graph-homomorphism

To avoid this problem, Otsuka and Saigo (2022) propose to combine the DAG and string diagram formalisms and define abstraction over both levels. As for abstraction of causal graphs, they require that a target "macro" graph $H = (\boldsymbol{V}_H, \boldsymbol{E}_H)$ be graph homomorphic to an original "micro" causal graph $G = (\boldsymbol{V}_G, \boldsymbol{E}_G)$, i.e., there is a mapping $\phi : \boldsymbol{V}_G \to \boldsymbol{V}_H$ such that if $X \to Y \in \boldsymbol{E}_G$ then $\phi(X) \to \phi(Y) \in \boldsymbol{E}_H$. This ensures the structural consistency (the first desideratum in the above list) between G and H. The graph homomorphism ϕ then induces an abstraction of string diagrams as a functor $\Phi : \mathsf{Syn}_G \to \mathsf{Syn}_H$, which sends an object (string) Y in Syn_G to object $\phi(Y)$ in Syn_H, and boxes:



(6)

where $Z_1 \dots Z_l \in \mathrm{PA}(\phi(Y)) \setminus \phi(\mathrm{PA}(Y))$ (note that the right box is indeed a morphism in Syn_H).

With this setup, a macro model functor $F' : \mathsf{Syn}_H \to \mathsf{FinStoch}$ is said to be a Φ-*abstraction* of a micro model $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$ if there is a natural transformation $\alpha : F \Rightarrow F'\Phi$, so that for any morphism $f : X \to Y$ in Syn_G the following diagram commutes:

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ {\scriptstyle \alpha_X}\big\downarrow & & \big\downarrow{\scriptstyle \alpha_Y} \\ F'\Phi(X) & \xrightarrow{F'\Phi(f)} & F'\Phi(Y) \end{array} \tag{7}$$

The difference from (5) is that the lower half now represents the stochastic transition in macro *graph* H. The commutativity thus ensures the probabilistic consistency (the second desideratum) between the micro causal model F based on DAG G and the macro model F' based on another DAG H. Otsuka and Saigo (2022, theorem 4) also show that the Φ-abstraction satisfies the interventional consistency, i.e., for any intervention on a macro-level variable there is a corresponding intervention on a set of micro variables such that these two interventions yield consistent post-intervention distributions.

Fig 4 illustrates the procedure of Φ-abstraction with a simple example, where two tips Y, Z of a fork $Y \leftarrow X \to Z$ are merged into one variables W. The middle

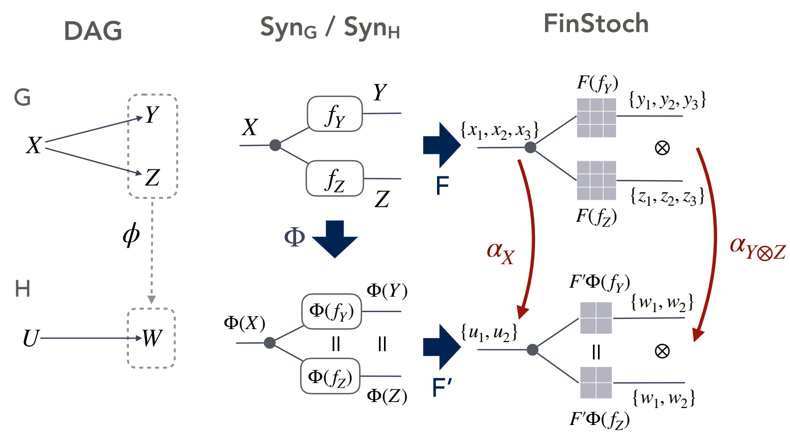


Figure 4: An example of Φ-abstraction, adapted from Otsuka and Saigo (2022). Here the causal flow goes from left to right. The graph homomorphism ϕ on the left column merges two effects Y, Z in DAG G into single variable W. The middle column shows how the induced functor $\Phi : \mathsf{Syn}_G \to \mathsf{Syn}_H$ operates on a string diagram in Syn_G. The natural transformation (red arrows) in the right column connects two models F, F' in the category $\mathsf{FinStoch}$.

column is string diagram representations of the corresponding DAGs on the left. Although the diagram below obtained from the abstraction functor Φ preserves the fork structure of the original diagram (above), the two branches are identical. Causal models F and F' interpret these string diagrams in $\mathsf{FinStoch}$(right column). Here "macro" variables X, Y, Z each have three values, while "micro" variables U, W have two. The morphisms $F(f_Y), F(f_Z)$ are then 3×3 stochastic matrices, while $F'\Phi(f_Y) = F'\Phi(f_Z)$ is 2×2. The Φ-abstraction in this case consists of a 2×3 matrix α_X and a $(2 \times 2) \times (3 \times 3)$ matrix $\alpha_{Y \otimes Z}$ that make the following diagram commutes:

$$
\begin{array}{ccccc}
F(X) & \xrightarrow{F(cp_X)} & F(X) \otimes F(X) & \xrightarrow{F(f_Y)\otimes F(f_Z)} & F(Y) \otimes F(Z) \\
{\scriptstyle \alpha_X}\downarrow & & & & \downarrow{\scriptstyle \alpha_{Y\otimes Z}} \\
F'\Phi(X) & \xrightarrow{F'(cp_{\Phi(X)})} & F'\Phi(X) \otimes F'\Phi(X) & \xrightarrow{F'\Phi(f_Y)\otimes F'\Phi(f_Z)} & F'\Phi(Y) \otimes F'\Phi(Z) \\
\| & & \| & & \| \\
F'(U) & \xrightarrow{F'(cp_U)} & F'(U) \otimes F'(U) & \xrightarrow{F'(f_W)\otimes F'(f_W)} & F'(W) \otimes F'(W).
\end{array}
\tag{8}
$$

In the above Fig. 4, the abstraction functor Φ replicates the fork structure in Syn_H. This construction is legitimate despite the lack of of a fork in DAG H, because the corresponding free category Syn_H is equipped with a copier.

Moreover, the result of the abstraction carries over to DAG H. The abstracted morphism $F'\Phi(f_Y) = F'\Phi(f_Z)$ that makes the above diagram (8) commutative is *ipso facto* the probabilistic interpretation $F'(f_W)$ of the morphism $f_W : U \to W$. This stochastic matrix, in turn, gives conditional probabilities $P(U|W)$ in DAG H which is consistent with $P(Y, Z|X)$ in the micro model F based on DAG G. Hence though the fork structure remains in the target string diagram Syn_H, its causal model functor F' that constitutes the Φ-abstraction can be interpreted as a macro level causal model on the DAG H that does not have the fork.

5 Conclusion

This paper reviewed the category-theoretic approach to causal modeling pioneered by Jacobs et al. (2019), and explored its philosophical as well as methodological implications. The categorical approach represents a causal structure as a diagrammatic network of mechanisms (box) connected via processes (wires), and defines a causal model as a functor that assigns a probabilistic interpretation to the diagram. This alternative perspective makes it clear the logical connection between the Markov condition and the modularity condition, and their dependence on the existence of a particular process called copier. The categorical approach also offers a natural method for abstracting causal models by the notion of natural transformation, combined with graph homomorphism.

Although the presentation in this paper focused on discrete causal models, we believe that it can be extended to continuous cases by considering functors to a more general category of measurable Markov kernels Stoch or its subcategory $\mathsf{BorelStoch}$ consisting of standard Borel spaces (Fritz, 2020). Another pending issue is the extension of the Φ-abstraction discussed in Sec. 4.2. Although it enables us to merge two parallel processes or forks as shown in Fig. 4, it cannot be used to collapse a cause-effect relationship $X \to Y$ into a single variable, because graph homomorphism then requires a self-loop at the codomain, making the graph no longer a DAG. The extension of the abstraction procedure to handle such cases will be a task for future work.

References

Abramsky S, Coecke B (2004) A categorical semantics of quantum protocols. In: Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004., pp 415–425

Beckers S, Halpern JY (2019) Abstracting causal models. AAAI 33(01):2678–2685

Beckers S, Eberhardt F, Halpern JY (2020) Approximate causal abstractions. In: Adams RP, Gogate V (eds) Proceedings of the 35th Uncertainty in Artificial Intelligence Conference, vol 115, pp 606–615

Cartwright N (1999) The Dappled World. Cambridge University Press

Cartwright N (2007) Hunting Causes and Using Them. Cambridge University Press

Chalupka K, Perona P, Eberhardt F (2014) Visual causal feature learning. In: Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence, pp 181–190

Chalupka K, Eberhardt F, Perona P (2016) Multi-Level Cause-Effect systems. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, vol 51, pp 361–369

Coecke B, Kissinger A (2017) Picturing Quantum Processes. Cambridge University Press

Dowe P (2000) Physical Causation. Cambridge University Press

Fritz T (2020) A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. Adv Math 370:107239

Hausman DM, Woodward J (1999) Independence, invariance and the causal markov condition. British Journal for the Philosophy of Science 50:521–583

Iwasaki Y, Simon HA (1994) Causality and model abstraction. Artif Intell 67(1):143–194

Jacobs B (2022) Structured probabilistic reasoning. Unpublished manuscript

Jacobs B, Kissinger A, Zanasi F (2019) Causal inference by string diagram surgery. In: Foundations of Software Science and Computation Structures, Springer International Publishing, pp 313–329

Machamer P, Darden L, Craver CF (2000) Thinking about Mechanisms. Philos Sci 67(1):1 25

Morgan SL, Winship C (2007) Counterfactuals and Causal Inference. Cambridge University Press p 524

Otsuka J, Saigo H (2022) On the equivalence of causal models: A Category-Theoretic approach. In: Proceedings of the First Conference on Causal Learning and Reasoning, pp 634–646

Pearl J (2000) Causality: Models, Reasoning, and Inference. Cambridge University Press

Peters J, Janzing D, Schölkopf B (2017) Elements of causal inference: foundations and learning algorithms. The MIT Press

Rischel EF (2020) CategoryTheory of causal models. PhD thesis, University of Copenhagen

Rischel EF, Weichwald S (2021) Compositional abstraction error and a category of causal models. In: de Campos C, Maathuis MH (eds) Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, PMLR, Proceedings of Machine Learning Research, vol 161, pp 1013–1023

Rubenstein PK, Weichwald S, Bongers S, Mooij JM, Janzing D, Grosse-Wentrup M, Schölkopf B (2017) Causal consistency of structural equation models. In: Elidan G, Kersting K, Ihler AT (eds) Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence

Salmon WC (1984) Scientific Explanation and the Causal Structure of the World. Princeton University Press

Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, Bengio Y (2021) Toward causal representation learning. Proc IEEE 109(5):612–634

Spirtes P, Glymour C, Scheines R (1993) Causation, Prediction, and Search. The MIT Press

Zennaro FM (2022) Abstraction between structural causal models: A review of definitions and properties 2207.08603