Behaviormetrika
https://doi.org/10.1007/s41237-023-00197-z

ORIGINAL PAPER

Process theory of causality: a category-theoretic perspective

Jun Otsuka[1,3] · **Hayato Saigo**[2]

Received: 19 September 2022 / Accepted: 25 March 2023

Abstract
This article presents an overview of the category-theoretical approach to causal modeling, as introduced by Jacobs et al. (2019), and describes some of its conceptual and methodological implications. Categorical formalism emphasizes causality as a process wherein a causal system is represented as a network of connected mechanisms. We demonstrate that this alternative perspective provides novel insights into the long-standing issue regarding the validity of the Markov condition, as well as formal mapping between micro-level and abstracted macro-level causal models.

Keywords Causal models · Symmetric monoidal category · String diagram · Markov condition · Abstraction · Causal representation learning

1 Introduction

At present, graphical modeling is the standard toolkit for studying causality and determining causal relationships from observed data (Spirtes et al. 1993; Pearl 2000). In this approach, a typical causal model $\mathcal{M} = (G, P)$ consists of a directed acyclic graph (DAG) G over a set of variables and a probability distribution P, wherein the graph $G = (\boldsymbol{V}, \boldsymbol{E})$ is a pair of a set $\boldsymbol{V}$ of variables and a set $\boldsymbol{E} \subset \boldsymbol{V} \times \boldsymbol{V}$ of the edges between them. The variables designate the properties or states of units or objects, such as the diets or blood pressures of patients. The existence of an edge

Communicated by Shohei Shimizu.

✉ Jun Otsuka
jotsuka@bun.kyoto-u.ac.jp

Hayato Saigo
harmoniahayato@gmail.com

[1] Department of Philosophy, Kyoto University, Kyoto, Japan

[2] Department of Frontier Bioscience, Nagahama Institute of Bio-Science and Technology, Nagahama, Japan

[3] Causal Inference Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

Published online: 11 April 2023

from one variable to another indicates that the state of the latter is causally dependent on that of the former such that an intervention in the former results in a change in the latter. Thus, within this framework, causality is understood as the relationship between events, where the events are designated by variables that assume particular values. For example, BloodPresure = high indicates an event wherein the blood pressure of a given patient is high, and a causal question asks whether such an event exhibits a systematic relationship with other events, such as diet or other medical conditions.

The event-centered view dates back to British Empiricism, particularly the work of David Hume, who considered inductive reasoning as an inference from one type of event to another. For Hume, this task was equivalent to establishing a causal relationship between events, which he believed could not be warranted by logic or experience. Studies on contemporary statistics and machine learning have attempted to address this skepticism by introducing various empirical and theoretical assumptions that enable an algorithmic identification of causal relationships from observed data (Morgan and Winship 2007; Peters et al. 2017). However, the basic conceptual framework remains the same: a causal system is considered as a constellation of events/variables that manifest regular patterns.

Several philosophers have proposed an alternative conception of causality, which features the aspects of processes (Salmon 1984; Dowe 2000) or mechanisms (Machamer et al. 2000; Cartwright 2007). According to this perspective, causality is best understood as a process that transmits influence from one event to another or a mechanism that produces an outcome by taking certain inputs. For example, a metabolic process may be considered a mechanism that "generates" blood pressure (among other factors) in response to, for example, a dietary practice.[1]

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 can provide novel insight into certain problems regarding the causal Markov condition and abstraction. Based on the seminal work of Jacobs et al. (2019), Sect. 2 presents the categorical formalization of discrete causal models with finite variables. We demonstrate how causal DAGs translate into string diagrams and that a functorial mapping of diagrams yields causal models. Our approach prioritizes clarity over theoretical rigor and proceeds via examples rather than mathematical proofs so that the reader can grasp the core concept without familiarity with category theory. In Sect. 3, we investigate the problem of the Markov condition from a categorical perspective and point out that the validity of this condition is dependent on the existence of a special mechanism known as the *copier*, which duplicates a causal process without disturbing it. Section 4 presents the problem of abstracting a causal model by coarsening its variables. The challenge of abstraction involves mapping a "low-level" micro model to a "high-level" macro model consistently. We demonstrate that this mapping can be produced via a category-theoretic notion of *natural transformation* between two causal

[1] Of course, these aspects of causality are not incompatible: for example, Pearl (2000) emphasized the mechanistic interpretation of structural equations in the graphical approach. We thank an anonymous reviewer for pointing this out.

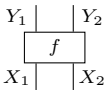
models/functors. We conclude that the category-theoretic approach offers a novel perspective on and solutions to certain issues that have resisted successful formal treatment in the conventional DAG formalism.

2 Process theory of causality

Although the event-centered view of causality is naturally represented in graphical modeling, the process-centered view can be formalized using process theory, which has mainly been developed in categorical quantum mechanics and computer science (e.g., Abramsky and Coecke 2004; Coecke and Kissinger 2017). In this section, we briefly review the application of process theory to causal modeling, as introduced by Jacobs et al. (2019).

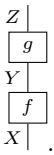
2.1 Translation of a DAG into a string diagram

Process theory conceptualizes a process as a system of combined mechanisms that exchange their products with one another. Each mechanism, which is commonly represented by a box, has definite types of inputs and outputs, which are represented by wires. An example of a mechanism that takes two inputs X_1, X_2 and returns two outputs Y_1, Y_2 is as follows:



Unless mentioned otherwise, a causal process flows from bottom to top.

Given two boxes f and g, if the output type of f matches the input type of g, these two boxes can be combined vertically using the matching wire, as follows:



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

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

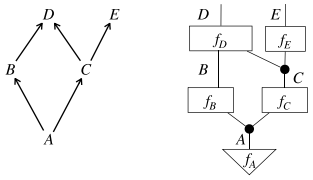
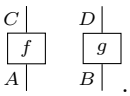


Fig. 1 Translation of a DAG (left) into a string diagram (right)

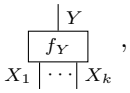


.

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

A system created by combining multiple mechanisms using vertical and parallel compositions is known as a *string diagram*. A whole string diagram can also be considered a large process that takes combined inputs and emits combined outputs.

In the context of causal modeling, a diagram serves as a causal graph that describes the topological features (i.e., the connectedness) of a causal system. The wires in a string diagram correspond to the variables. For each variable $Y \in \boldsymbol{V}$, there is a box of the following form:



, (1)

where $X_1, \cdots, X_k \in \mathrm{PA}(Y)$ are the parents of Y. The box intuitively represents a "generating mechanism" of Y that takes $\mathrm{PA}(Y)$ as the input, and thus, multiple edges pointing to a variable are summarized by one box. Furthermore, we assume that an exogenous variable (with $\mathrm{PA}(Y) = \emptyset$) has its own "state" with no input, which is depicted by a triangle. A matching string diagram can be created by combining these boxes and wires in accordance with a given DAG, as illustrated in Fig. 1. Note that a string diagram provides a somewhat "flipped" image of the graph, wherein the nodes are replaced with wires and the edges with boxes.

A 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 required when a fork $X \leftarrow Y \rightarrow Z$ exists in the graph. From a process perspective, this means that the product Y is used twice: as an input to (the generating mechanism of) X and an input to Z. Such an operation is taken for granted in causal graphs, but not in string diagrams, and must be explicitly considered as an independent process because duplication is not always possible. For example, in quantum mechanics, one state cannot be copied without being disturbed. In Sect. 3, we discuss how the existence of the copier is also crucial for the validity of the causal Markov condition.

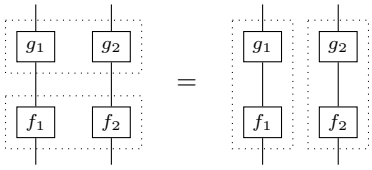
String diagrams can be formally described by the language of the symmetric monoidal category. The wires and boxes in a string diagram are objects and morphisms (arrows), respectively, in this category. A vertical composition of boxes corresponds to the composition of morphisms with a matching codomain/domain; 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 determined by the binary associative operations of objects and morphisms:

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

where $\mathrm{ob}(\mathcal{C})$ is a class of objects and $\mathcal{C}(A, B)$ is a 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, and 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 diagrammatic presentation, this simply means that the two means of composing processes



yield the same diagram.

Within the aforementioned categorical background, Jacobs et al. (2019) introduced a free category (also known as the *free CDU category*, wherein CDU stands for copy, discard, and uniform) over a pair of generating sets of objects and morphisms.[2]. In particular, a causal string category Syn_G is constructed from a DAG $G = (\boldsymbol{V}, \boldsymbol{E})$ using its variable set $\boldsymbol{V}$ as the generating set of objects and the set of boxes with the form (1) as the generating set of morphisms. Thus, Syn_G contains everything that can be obtained by simply combining these wires and boxes (as well

[2] This was later generalized by Fritz (2020) with the name *Markov category*.

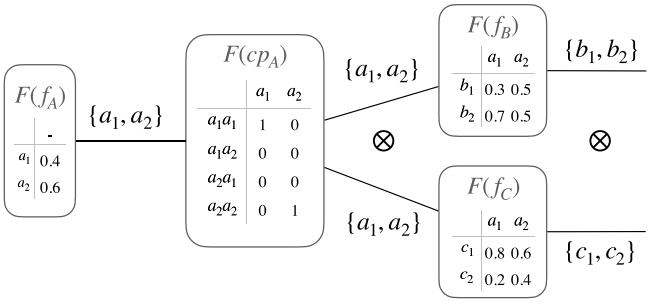


Fig. 2 Example of a functorial assignment of values and (conditional) probabilities to a string diagram. The causal flow is from left to right. The structure interprets the bottom part of the string diagram in Fig. 1

as other special units, such as copiers, discards, and units). This includes the string diagram in Fig. 1 as well as 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 free CDU category defined previously. This process is achieved using a functor, which is a systematic mapping from one category to another; in this case, from Syn_G to another CDU category with an appropriate structure (Jacobs et al. 2019; Fritz 2020). The target category for discrete causal models with variables that have only finite values is FinStoch, wherein the objects are finite sets and morphisms $f : X \to Y$ are $|Y| \times |X|$-dimensional stochastic matrices, i.e., matrices of nonnegative numbers in which the sum of each column is 1. A functor assigns each wire of Syn_G a finite set (representing the values of the corresponding variable) and each box a stochastic matrix (representing the conditional probabilities of an effect given its causes, which are also known as Markov kernels). Moreover, a state (a triangle with no input) of an exogenous wire/variable X is mapped to a morphism from object 1 of FinStoch. This morphism is a $|X| \times 1$ stochastic matrix or vector, and thus, yields the marginal distribution $P(X)$ of X.

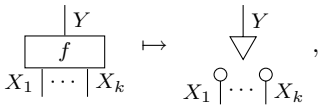
Figure 2 depicts probabilities assigned by a causal model functor F to the bottom half of the string diagram shown in Fig. 1. In this case, each variable/wire is assumed to have two values, and thus, mapped to the two-element sets $\{a_1, a_2\}$, $\{b_1, b_2\}$, and $\{c_1, c_2\}$. The leftmost box $F(f_A)$ provides a marginal distribution $P(A)$ in the 2×1 vector format. $F(cp_A)$ interprets the copier using a $(2 \times 2) \times 2$ matrix that effectively "duplicates" $P(A)$ to yield $P(A \times A)$. This is subsequently fed into $F(f_B)$ and $F(f_C)$, which are 2×2 matrices that represent the conditional distributions $P(B|A)$ and $P(C|A)$, respectively. Overall, the functor yields the joint probability distribution $P(A, B, C)$ that satisfies the Markov condition with the DAG $B \leftarrow A \rightarrow C$.[3]

[3] In string diagrams, only the wires that extend to the end are assumed to be observed. Hence, to obtain a joint distribution $P(A, B, C)$, A must be branched once more and run to the end. However, in this study, we ignore this convention and assume that all wires in a string diagram are observed.

A different functor $F' : \mathsf{Syn}_G \to \mathsf{FinStoch}$ leads to a different probability assignment, possibly with varying numbers of variable values. 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

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



while all other boxes and wires remain 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 define similar endofunctors. A post-intervention distribution is obtained by combining an intervention functor with a causal model functor such that $F \cdot \mathsf{cut}_Y : \mathsf{Syn}_G \to \mathsf{FinStoch}$.

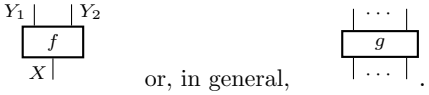
3 Markov condition

Thus far, we have reviewed the categorical formalization of causal models in Jacobs et al. (2019) as a formal representation of the process-oriented perspective of causality. The advantage of considering this alternative perspective is that it provides insight into issues that resist proper theoretical handling in the conventional DAG formalism. Jacobs et al. (2019) demonstrated that the identifiability of intervention outcomes can easily be determined via the diagrammatic operation known as comb disintegration. In this and the following sections, we discuss two other issues: the Markov condition and abstraction of causal models.

A causal model (G, P) with a directed graph $G = (\boldsymbol{V}, \boldsymbol{E})$ satisfies the global Markov condition when the joint distribution P is factorized as $P(\boldsymbol{V}) = \prod_{X \in \boldsymbol{V}} P(X|\mathrm{PA}(X))$, where $\mathrm{PA}(X)$ denotes the parents of X in G. This implies the local Markov condition in which each variable X is independent of its non-descendants given its parents $\mathrm{PA}(X)$. As noted at the end of the previous section, discrete causal model (G, P) that satisfies the Markov condition corresponds one-to-one with functor $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$. However, this does not imply the

equivalence of the diagrammatic and graph-theoretic formalizations. The former can deal with a broader range of causal structures, including non-Markovian structures.

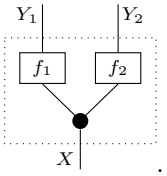
To see this, note that the aforementioned procedure for constructing a string diagram from a causal graph is based on the assumption that each variable/wire has its own generating mechanism, represented by a box with only one output. However, in general, in process theory (or the symmetric monoidal category), boxes may have multiple outputs, such as



(2)

As 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 graph-theoretical counterparts (Jacobs 2021).

Note that the left box f in (2) is *not* equivalent to fork $Y_1 \leftarrow X \rightarrow Y_2$; if it were a fork, the Markov condition would entail $Y_1 \perp Y_2|X$, but nothing in the diagrammatic representation enforces this independent 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. The independence can be assured with the use of a copier:



(3)

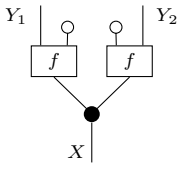
This is the correct diagrammatic rendition of the fork $Y_1 \leftarrow X \rightarrow Y_2$ in a causal DAG, which makes Y_1 and Y_2 independent given X in any functorial (probabilistic) interpretation of this diagram. Furthermore, because every box in this diagram has only one output, it can be constructed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

Alternatively, the causal Markov condition can be understood as the requirement that every multi-output process, as in (2), must be a disguised dashed box, as in (3), and 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 using 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 appropriately named the *modularity condition*. This further implies that each endogenous variable/wire has its own box/mechanism that is distinct from the other boxes in the diagram. In previous attempts to prove the Markov condition, the modularity was defined as the independent manipulability of each variable, which leaves the structural equations of the other variables intact (Hausman and Woodward 1999). However, this definition does not readily extend to probabilistic cases, which have been the touchstone case for the

validity of the Markov condition (Cartwright 2007). The diagrammatic condition in (3) better captures the underlying concept of manipulability: "causes are as it were levers that can be used to manipulate their effects" (Hausman and Woodward 1999, p. 533), and the manipulability in this sense does imply the common cause principle, the central as well as controversial part of the Markov condition, whereby multiple effects of the same cause become independent of one another given their common causes, provided that they are not causes or effects of one another.

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 argued that the Markov condition fails when a cause operates probabilistically and illustrated her claim using a hypothetical chemical factory that generates products Y_1 and pollutants Y_2 with certain probabilities such that Y_1 and Y_2 do not become independent, even conditionally on the operation X of the factory (Cartwright 2007, p. 107). This factory is equivalent to the process f in (2), and Cartwright's claim is that it is not decomposable as in (3), because the chemical products and pollutants are assumed to be generated via the same mechanism.

Her argument can be paraphrased using diagrams: if f in (2) is equivalent to (3), it can also be rewritten as



(4)

where the empty circles are operators that "discard" each of the two outputs Y_1 and 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 does appear to be a rather strong assumption that is unlikely to hold in situations such as the example of Cartwright.

Cartwright's hypothetical chemical factory is an example of *interactive forks*, as labeled by Salmon (1980), wherein "the change in each process is produced by the interaction with the other process" (p. 12). In Cartwright's example, the processes that generate products Y_1 and pollutants Y_2 supposedly interact with each other through chemical reactions. Salmon contrasted this type of causal mechanism with *conjunctive forks*, wherein "the dependency [among effects] arises, not because of any physical interaction [...] but because of special background conditions" (p. 9). His example comprised two identical term papers submitted by different students, independently plagiarizing a common source. Another more scientific example is a pleiotropic gene with multiple phenotypic effects, such as the abnormal β-globin gene, which is responsible for both sickle cell disease and malarial resistance. In both examples, the common causes serve as shared "background conditions," by being copied (by the students or DNA transcription) each time they produce effects.

Salmon claimed that the common cause principle holds with conjunctive forks but not with interactive forks.

This claim is corroborated by the categorical framework. The difference between two types of causal forks can be formally represented by the presence or absence of a copier. In a conjunctive fork, the processes that generate each of its effects operate independently on copied inputs ("background conditions"), as in (3). In such cases, the dependency between the two effects vanishes when conditioned on their common cause. However, if the fork is interactive, the production is not mediated by a copier, and thus, the common cause principle does not necessarily hold as discussed above. In this manner, the copier plays the central role in the common cause principle and the Markov condition.

4 Abstraction of causal models

The next problem we consider is that of abstracting causal models. Causal systems can be described at different levels of granularity, and the determination of appropriate macro-level causal features from micro-level measurements (such as gene expression data or image pixels) is a major challenge in machine learning and scientific inquiries (Iwasaki and Simon 1994; Chalupka et al. 2014, 2016; Schölkopf et al. 2021). The assumption of coarsening is that models at different levels must be consistently related despite having different sets of variables and edges for them to be considered to model the same phenomenon. Recent studies have proposed formal conditions of such an abstraction procedure that maps the 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 a lower resolution, or both (Zennaro 2022). In any case, this mapping must be consistent in three essential aspects of causal models for the resulting model to be considered an abstraction of the original model:

1 Structural: The causal relationships of the low-level model must be preserved. In particular, if an edge exists 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 regarding external interventions.

These desiderata together require that the abstraction procedure commute with various operations in/on a causal model. For example, probabilistic consistency requires that the probability of an effect calculated in the micro model "match" that

of its macro counterpart (Rubenstein et al. 2017; Rischel 2020; Rischel and Weichwald 2021). In the following sections, we demonstrate that the category-theoretic formulation provides a natural micro–macro translation that fulfills all of these requirements.

4.1 Abstraction in monoidal category

We begin with value reduction, of which two types are possible. The first is deterministic transformation or *supervenience*, which merges multiple values of one variable into fewer values with a lower resolution. In cases of discrete variables, such a map is obtained via a rank-deficient stochastic matrix with entries of 1 or 0. The second type is stochastic, which simply maps one variable to another using any stochastic matrix with a size that matches the number of values of the source and target variables. The categorical approach handles both types in the same manner using the concept of natural transformation.

Suppose that we are given a causal model $F : \mathsf{Syn}_G \to \mathsf{FinStoch}$. An abstracted model that merges several values of its variables is represented 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, the abstraction is a mapping between functors $F \Rightarrow F'$ that fulfills the consistency requirements listed previously. In category theory, such a mapping is known as a *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 $\mathsf{FinStoch}$ (stochastic matrices) that make the following diagram commute for any morphism $f : X \to Y$ in Syn_G:

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

where the upper half represents a stochastic transition along the causal edge $f : X \to Y$ according to the original model F, and the lower half represents the corresponding transition in the coarse-grained model F'. These are the stochastic matrices of dimensions $|F(Y)| \times |F(X)|$ and $|F'(Y)| \times |F'(X)|$, respectively. In contrast, the vertical arrows α_X and α_Y relate these causal flows in a consistent manner. These are also stochastic matrices, the entries of which are either 1 or 0 in the case of deterministic translation (i.e., merging of values). The commutativity of the diagram indicates that the coarsening of $\alpha_X : F(X) \to F'(X)$ is consistent at every step of the causal flow in the sense that the same marginal distribution is obtained regardless of whether the causal path in the original model is followed and the effect is transformed (clockwise path) or the cause is first transformed and its effect is then derived in the coarse-grained model (counterclockwise path). Thus, the existence of a natural transformation between the two models/functors F and F' warrants probabilistic consistency.

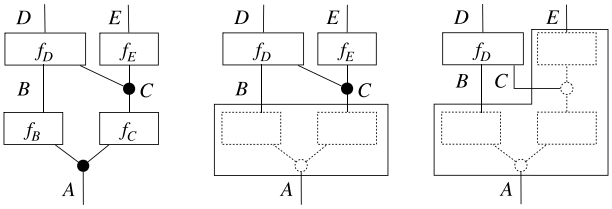


Fig. 3 "Abstraction" with string diagrams. In symmetric monoidal categories, objects (wires) and morphisms (boxes) can be combined to form a joint process. The string diagram in the middle combines a copier and two parallel processes f_B and f_C into one process. The inverse L-shaped box on the right further encompasses another copier and f_E, thereby constituting a process with three outputs B, C, and E

In general, the determination of an abstraction between two candidate models is a non-trivial task. However, in deterministic abstraction, there is a necessary and sufficient condition for the existence of a transformation (Otsuka and Saigo 2022). This condition is called *causal homogeneity*, which intuitively requires that the micro values to be merged into the same macro value must have homogeneous causal effects modulo groups of the effect variable. For further details, refer to Otsuka and Saigo (2022). Alternatively, Rischel (2020) and Rischel and Weichwald (2021) proposed the use of KL-divergence to measure the non-commutativity of abstraction when an exact match between two models does not hold, which is expected in empirical measurements.

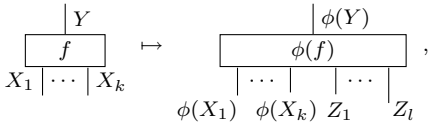
We now move on to the problem of variable reduction, wherein two or more variables in a micro model are merged into one variable in a macro model. In a way, this type of merging is already built into the monoidal category as vertical or horizontal compositions in a string diagrams. Recall that Syn_G, as a free symmetric monoidal category, contains appropriate compositions of the generating objects and morphisms. Such combined objects or morphisms can be considered "abstractions" of its components. For example, Fig. 3 depicts the progressive procedures by which the components are combined to form larger processes, which can be considered abstractions of their constituting processes. In this sense, the horizontal and vertical compositions of the string diagram provide a means of variable reduction. The functorial property of a causal model then takes care of both probabilistic and interventional consistencies. In particular, the probabilistic interpretation of the merged processes can be calculated from those of their constituents.

However, categorical/monoidal compositions cannot be considered complete abstractions by themselves. Abstraction is expected to consolidate information as well as discard or forget some of it. Composition may serve the former but not the latter purpose, as the composed boxes or wires retain all details as their components. Moreover, this procedure does not allow one to compare two causal graphs. The boxes that result from compositions may have multiple outputs, in which case there may be no graph-theoretic counterpart with a visible abstract relationship to the original graph (Sect. 3). For example, no causal graph corresponding to the middle

and right string diagrams in Fig. 3 exists that preserves the cause–effect relationships in the original causal graph (Fig. 1). A different approach must be adopted to understand abstraction in the conventional graphical formalism.

4.2 Abstraction via graph homomorphism

To avoid the aforementioned problem, Otsuka and Saigo (2022) proposed a combination of the DAG and string diagram formalisms and defined the abstraction over both levels. The abstraction of causal graphs requires that a target "macro" graph $H = (\boldsymbol{V}_H, \boldsymbol{E}_H)$ correspond to an original "micro" causal graph $G = (\boldsymbol{V}_G, \boldsymbol{E}_G)$. The correspondence can be spelled out by a graph homomorphism $\phi : \boldsymbol{V}_G \to \boldsymbol{V}_H$ such that if $X \to Y \in \boldsymbol{E}_G$ then $\phi(X) \to \phi(Y) \in \boldsymbol{E}_H$. This ensures structural consistency (the first requirement in the aforementioned list) between G and H. The graph homomorphism ϕ induces an abstraction of string diagrams as a functor $\boldsymbol{\Phi} : \mathsf{Syn}_G \to \mathsf{Syn}_H$, which sends an object (string) Y in Syn_G to object $\phi(Y)$ in Syn_H, and boxes



(6)

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

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

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

The difference from (5) is that the lower half represents the stochastic transition in the macro *graph* H. This commutativity ensures probabilistic consistency (the second requirement) between the micro causal model F based on the DAG G and the macro model F' based on another DAG H. Otsuka and Saigo (2022, Theorem 4) also demonstrated that the $\boldsymbol{\Phi}$-abstraction satisfies 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.

Figure 4 depicts the $\boldsymbol{\Phi}$-abstraction procedure using a simple example, where the two tips Y and Z of a fork $Y \leftarrow X \to Z$ are merged into one variable W with fewer values. The middle column shows the string diagram representations of the corresponding DAGs on the left side. Although the lower diagram, which is obtained from the

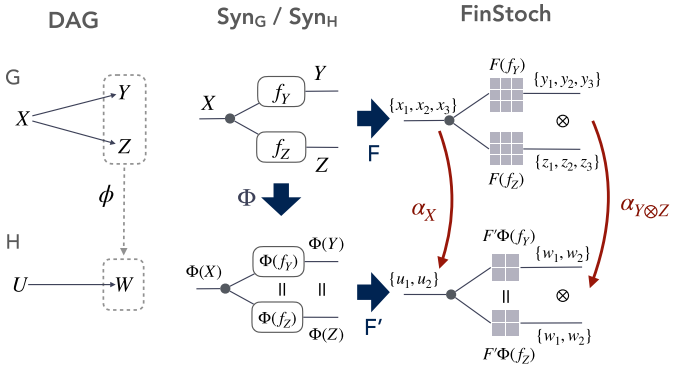


Fig. 4 Example of the reduction in both variables and values via $\boldsymbol{\Phi}$-abstraction, adapted from Otsuka and Saigo (2022). The causal flow is from left to right. The graph homomorphism ϕ on the left column merges two effects Y and Z in the DAG G into a single variable W. The middle column indicates how the induced functor $\boldsymbol{\Phi} : \mathsf{Syn}_G \to \mathsf{Syn}_H$ operates on a string diagram in Syn_G. The natural transformation (curved arrows) in the right column connects two models F and F' in the category FinStoch

abstraction functor $\boldsymbol{\Phi}$, preserves the fork structure of the original diagram (above), the two branches are identical. The causal models F and F' interpret the string diagrams in FinStoch(right column). In this case, the "micro" variables X, Y, and Z each have three values, whereas the "macro" variables U and W have two. Thus, the morphisms $F(f_Y)$ and $F(f_Z)$ are 3×3 stochastic matrices, whereas $F'\boldsymbol{\Phi}(f_Y) = F'\boldsymbol{\Phi}(f_Z)$ is 2×2. The $\boldsymbol{\Phi}$ -abstraction 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 commute:

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

In Fig. 4, the abstraction functor $\boldsymbol{\Phi}$ replicates the fork structure in Syn_H. This construction is legitimate despite the lack of a fork in the DAG H, because the corresponding free category Syn_H is equipped with a copier. Moreover, the result of the abstraction is carried over to the DAG H. The abstracted morphism $F'\boldsymbol{\Phi}(f_Y) = F'\boldsymbol{\Phi}(f_Z)$ 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 the DAG H, which is consistent with $P(Y, Z|X)$ in the micro model F based on the DAG G. Hence, although

the fork structure remains in the target string diagram Syn_H, its causal model functor F', which constitutes the Φ-abstraction, can be interpreted as a macro-level causal model on the DAG H, which does not have a fork.

5 Conclusions

This study has reviewed the category-theoretic approach to causal modeling pioneered by Jacobs et al. (2019) and investigated its philosophical and methodological implications. The categorical approach represents a causal structure as a diagrammatic network of mechanisms (boxes) connected via processes (wires), and defines a causal model as a functor that assigns a probabilistic interpretation to the diagram. This alternative perspective clarifies the logical connection between the Markov and modularity conditions and their dependence on the existence of a particular process known as the copier. Moreover, the categorical approach offers a natural method for abstracting causal models using the notion of natural transformation combined with graph homomorphism.

Although the approach in this study has focused on discrete causal models, it may 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 issue that needs further investigation is the extension of the Φ-abstraction, as discussed in Section 4.2. Although this procedure enables two parallel processes or forks to be merged, as illustrated in Fig. 4, it cannot be used to collapse a cause-effect relationship $X \to Y$ into a single variable, because the graph homomorphism in such a case requires a self-loop in the codomain, which results in the graph no longer being a DAG. Here, the concept of *ϕ-refinement*, which was recently proposed by Yin (2022), may be useful. However, a thorough examination of these issues is beyond the scope of the current study and remains a task for future research.

Acknowledgements The authors would like to thank Shohei Shimizu for his support and two anonymous reviewers for comments on earlier version of this paper.

Funding This work was supported by JSPS KAKENHI Grant Numbers 19K00270 (for JO), 19K03608 and 20H00001 (for HS).

Data availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

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, pp 415–425

Beckers S, Halpern JY (2019) Abstracting causal models. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, vol 33, pp 2678–2685
Beckers S, Eberhardt F, Halpern JY (2020) Approximate causal abstractions. In: Proceedings of the 35th Uncertainty in Artificial Intelligence Conference, vol 115, pp 606–615
Cartwright N (1999) The Dappled World. Cambridge University Press, Cambridge
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, Cambridge
Dowe P (2000) Physical Causation. Cambridge University Press, Cambridge
Fritz T (2020) A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. Adv Math. https://doi.org/10.1016/j.aim.2020.107239
Hausman DM, Woodward J (1999) Independence, invariance and the causal Markov condition. Br J Philos Sci 50:521–583
Iwasaki Y, Simon HA (1994) Causality and model abstraction. Artif Intell 67(1):143–194
Jacobs B (2021) Structured probabilistic reasoning. https://www.cs.ru.nl/B.Jacobs/PAPERS/ProbabilisticReasoning.pdf. Accessed 7 Feb 2023
Jacobs B, Kissinger A, Zanasi F (2019) Causal inference by string diagram surgery. 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
Otsuka J, Saigo H (2022) On the equivalence of causal models: a category-theoretic approach. Proc First Conf Causal Learn Reason PMLR 177:634–646
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) Category Theory of causal models. PhD thesis, University of Copenhagen
Rischel EF, Weichwald S (2021) Compositional abstraction error and a category of causal models. In: Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence, PMLR 161:1013–1023
Rubenstein PK, Weichwald S, Bongers S, Mooij JM, Janzing D, Grosse-Wentrup M, Schölkopf B (2017) Causal consistency of structural equation models. In: Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence
Salmon WC (1980) Causality: Production and Propagation. In: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 1980(2):49-69
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
Yin Y (2022) A graphical construction of free Markov categories. arXiv:2204.04920
Zennaro FM (2022) Abstraction between structural causal models: A review of definitions and properties. arXiv:2207.08603