Path Intervention for Path-specific Effects

Anonymous CVPR 2021 workshop submission

Paper ID none

Abstract

Path-specific effects in mediation analysis provides a useful tool for fairness analysis, which are mostly based on nested counterfactuals. However, the dictum "no causation without manipulation" implies that path-specific effects might be induced by certain interventions. This paper propose the path intervention inspired by information accounts of causality, corresponding intervention diagrams and π-formula for them are also developed. Comparing to the interventionist approach of [20] based on nested counterfactuals, our proposed path intervention method explicitly describes the manipulation in structural causal model(SCM) with simple information transferring interpretation and does not require the non-existing of recanting witness to identify path-specific effects, which could be an theoretical focus and facilitate communication with subject matter experts for mediation analysis.

1. Introduction

Causality has been one of the basic topics of philosophy since the time of Aristotle. Recently it has been claimed that an information accounts of causality, formally proposed and developed since [5] in philosophy, might be useful to the scientific problem of how we think about, and ultimately trace, causal linking, and so to causal inference and reasoning. Henceforth, from the perspective of causal modeling, is this philosophical production account of causality truly useful or just a rhetorical flourish?

To answer the above question, let's turn to one of seven tasks summarized by Judea Pearl in causal inference s[15]. Analyzing the relative strength of different pathways between a decision X and an outcome Y is a topic that has interested both scientists and practitioners across disciplines for many decades. Specifically, mediation analysis are useful concepts to address the problem [3, 19, 13]. More recently, path-specific effects are studied with approaches based on nested counterfactual [1, 23, 24, 11]. The dictum "no causation without manipulation" implies that interventions for path-specific effects should be proposed, and efforts are made by [20] based on single world intervention graph(SWIG) but still without a concise and formally defined intervention. A simple question to the nested counterfactual approach for path-specific effects would be "In Fig. 1, what effects of simply assigning the treatment $do(a')$ exclusively to the red causal path $\pi : A \to M \to Y$ while allowing all other causal paths to behave naturally?".

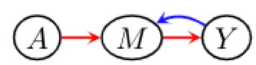

Figure 1. An example that the path-specific effect can be hard to define for nested counterfactual approach.

Fortunately, we found that the information accounts of causality can be very useful in developing semantic framework for path interventions. In particularly, for a structural causal model(SCM), we can decompose those associated causal mechanisms into two part: information transfer mechanisms and information process mechanisms. Then we proposed the novel path intervention as modification of information transfer mechanisms which directly assumes functional relationship between cross-world variables.

Comparing to the previous used approaches, our path-specific effects can be interpreted as simply making the treatment information $A = a'$ exclusively pass through the causal path π while allowing all other information outside from the π to behave naturally. We are not using nested counterfactual which may be problematic even in a simple example in Fig. 1, instead, we are using information view of **SCM**. Henceforth, we are able to develop many tools, such as intervention diagram and π-formula, for identification of path-specific effects. In particularly, these effects can be identified without satisfying the recanting witness criteria due to the difference in definition.

The remainder of this paper is organized as follows. Section 2 gives the preliminaries on information account of causality. Section 3 introduces our *path* intervention based on information view of SCM. Intervention diagrams are proposed in Section 4. Section 5 presents the π-formula for identification of path-specific effects induced by path intervention. Concluding remarks are offered in Section 6.

Notations. Fix a set of indices $V \equiv \{1, ..., n\}$. For each index $i \in V$ associate a random variable X_i with state space $\mathcal{X}_i$. Given $A \subseteq V$, we will denote subsets of random variables indexed by A as $X_A \in \mathcal{X}_A \equiv \prod_{i \in A} \mathcal{X}_i$. For notational conciseness we will sometimes use index sets to denote random variables themselves, using V and A to denote X_V and X_A, respectively, and similarly using lower case a to denote $x_A \in \mathcal{X}_A$. Similarly, by extension, we will also use V_A to denote X_A and V_i to denote X_i.

2. Information Account of Causality

Causality has been one of the basic topics of philosophy since the time of Aristotle. However during the last two decades or so, interest in causality has become very intense in the philosophy of science community, and a great variety of novel views on the subject have emerged and been developed. Among those novel views, it has been claimed that an informational account of causality, formally proposed and developed since [5] in philosophy, might be useful to the scientific problem of how we think about, and ultimately trace, causal linking, and so to causal inference and reasoning. An informational account of causality can be useful to help us reconstruct how science builds up understanding of the causal structure of the world. Broadly, we find mechanisms that help us grasp causal linking in a coarse-grained way. Then we can think in terms of causal linking in a more fine-grained way by thinking informationally. An informational account of causality may also give us the prospect of saying what causality is, in a way that is not tailored to the description of reality provided by a given discipline. And it carries the advantage over other causal metaphysics that it fares well with the applicability problem for other accounts of production (processes and mechanism)[8]. But from an application perspective, what benefits we can gain in causal modeling? Or it is just a rhetorical flourish?

At the centre of modern causal modeling theory lies the structural causal model (SCM) (also known as structural equation model) which makes graphical assumptions of the underlying data generating process (There many somewhat different formulations of SCM in literatures, e.g., [21, 14, 4, 17], among which we use the definition in [2] for this paper).

Definition 1 (SCM) *A structural causal model(SCM) $\mathcal{M}$ is a 4-tuple $(\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$, where*

- *$\mathbf{U}$ is a set of background variables, also called exogenous variables, that are determined by factors outside the model;*
- *$\mathbf{V}$ is a set $\{V_1, V_2, ..., V_n\}$ of variables, called endogenous, that are determined by other variables in the models- that is, variables in $\mathbf{U} \cup \mathbf{V}$;*
- *$\mathcal{F}$ is a set of functions $\{f_1, f_2, ..., f_n\}$ such that each f_i is a mapping from(the respective domains of) $U_i \cup Pa_i$ to V_i, where $U_i \subseteq \mathbf{U}$, $Pa_i \subseteq \mathbf{V} \setminus V_i$, and the entire set $\mathcal{F}$ forms a mapping from $\mathbf{U}$ to $\mathbf{V}$. That is, for $i \in V = \{1, ..., n\}$, each $f_i \in \mathcal{F}$ is such that*

$$v_i \leftarrow f_i(pa_i, u_i), \tag{1}$$

- *$P(\mathbf{U})$ is a probability function defined over the domain of $\mathbf{U}$.*

Mechanisms are activities and entities organized to produce some phenomenon which are represented by structural equations in an **SCM**. A structural model is Markovian if the exogenous parent set U_i, U_j are independent whenever $i \neq j$, and the associated causal graph for a Markovian SCM $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$ can be defined as follows:

Definition 2 (Causal Diagram (Markovian Models)) *$\mathcal{G}(\mathcal{M}) = (V, E)$ is said to be a* causal diagram *(of $\mathcal{M}$) if constructed as follows:*

- *add a vertex for every endogenous variable in the set V,*
- *add an edge $(V_j \rightarrow V_i)$ for every $i \in V$ if V_j appears as an argument of $f_i \in \mathcal{F}$ in the edge set E.*

Various difference-making and production accounts in philosophy have been taken account by Judea Pearl for developing the structural causal modeling framework. However, the information account of causality, one of six aspects of causality which can be classified into two groups of difference-making and production accounts [8], has been comparatively neglected. The view of causality as information transfer is first proposed in [5], further developed by [9, 8] and recently applied in neuroscience[6]. Inspired by this view of causality, we propose that the distribution of every endogenous variable V_i is decided by the information accepted from its input edges, the mechanism f_i decide how to process the accepted information and what information transferred on direct edges in the causal diagram. We then can reformulate SCM in the following:

Definition 3 (Information view of SCM) *Given a structural causal model(**SCM**) $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$, we reformulate the structural equations as*

$$\begin{cases} v_i \leftarrow f_i(e_{pa(i),i}, u_i) \\ e_{j,i} \leftarrow v_j \end{cases} \tag{2}$$

where $e_{j,i}$ represents the information on edge (j, i) received from its input node V_j.

Essentially, it can be interpreted as separating every causal mechanism f_i into two part — information process and information transfer. Mechanisms could be seen as information channels which is one centre idea of the informational accounts[5, 8]. Under this information transferring interpretation of SCM, both *do* intervention[12] and *info* intervention [7] are defined as following:

Definition 4 (*Do* and *info* intervention) *Given an SCM $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$ and any $I \subseteq V, v'_I \in \mathcal{V}_I$, the the* do intervention $do(v'_I)$ *can be defined as a modification of structural equations to*

$$\begin{cases} v_i \leftarrow f_i(e_{pa(i),i}, u_i) \; if \, i \notin I \; else \; v'_i \\ e_{j,i} \leftarrow v_j \end{cases} \tag{3}$$

while keeping everything else constant. And info intervention $\sigma(V_I = v'_I)$ *(or, in short, $\sigma(v'_I)$) maps $\mathcal{M}$ to the info-intervention model $\mathcal{M}^{\sigma(v'_I)}$ with a modified mechanism for every $i \in V$:*

$$\begin{cases} v_i \leftarrow f_i(e_{pa(i),i}, u_i) \\ e_{j,i} \leftarrow v_j \; if \, j \notin I \; else \; v'_j \end{cases} \tag{4}$$

while keeping everything else constant.

The different formulations of SCM reflect important epistemological distinctions about causality and lead to an alternative approach for mediation analysis. Comparing to the *do* intervention, the *info* intervention changes the information transferring instead of information processing mechanisms, which facilitates communication and theoretical focus as complementing causal modeling in terms of *do* operator. This motivates us to proposed a novel path intervention as modification of information transfer mechanisms.

It is also worth to mention that, broadly speaking, the information accounts of causality can also facilitate interpretation of existing widely used causal propositions. For example, regarding back-door/front-door criteria, the goal of which can be consistently considered as whether the observational information of a set of variables is enough or not to answer causal-effect estimation question, instead of conventional understanding such as controlling variables. Moreover, in general sense of information accounts, Pearl points out that questions in one layer of the causal hierarchy can only be answered when corresponding layer information are available[15, 2], and Scholköpf believes causal science will enable us act and decision with information from Lorenzian imagined space [21]. The recently proposed Mini-Turing test for AI — How can machines represent causal knowledge in a way that would enable them to access the necessary information swiftly, answer questions correctly, and do it with ease, as a human can? [18]. In summary, to build true intelligent machines, climb the ladder of data, information, knowledge and wisdom, we might need to incorporate the information accounts of causality into causal tasks.

In the following section, we a propose novel path intervention for path-specific effects based on information transfer view of SCM.

3. Path Intervention

Causal questions, such as what if we make something happen, can be formalized by using *do* intervention. However, the *do* intervention makes the intervention variable deterministic, by hypothetically forcing its value to a given constant. Henceforth, this causal notation is powerful but sometimes not enough for certain causal tasks, the well-known case is the nature directed(indirected) effect, which is defined by nested counterfactuals. However, the dictum "no causation without manipulation" suggests an interventionist to mediation analysis. Recently, an interventionist approach has already been proposed based on Single World Intervention Graphs [20] mediation analysis, but without clear and explicit mathematical form of what is the intervention for path-specific effects. Here, we are giving a explicit definition of path intervention based on SCM. Specifically, for a intervention variable A, the outcome variable of interest Y, and a path π from A to Y, the *path intervention* is defined as follows.

Definition 5 (Path intervention) *Given a Makovian SCM $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}))$, the path intervention $\pi(A = a')$ (or, in short, $\pi(a')$) along a causal path π from A to Y maps $\mathcal{M}$ to the path-intervention model $\mathcal{M}^{\pi(a')}$ with a modified causal mechanisms for every $i \in V, v_i^{\pi(a')}$(or in short v_i^{π}) and v_i are in corresponding counterfactual and factual domains, and u'_i is a realization of $\mathbf{U}'_i$ where $\mathbf{U}'$ is an i.i.d.[1] copy of $\mathbf{U}$:*

$$\begin{cases} v_i \leftarrow f_i(e_{pa(i),i}, u_i) \\ e_{j,i} \leftarrow v_j \\ v_i^{\pi} \leftarrow f_i(e'_{pa(i),i}, u'_i) \\ e'_{j,i} \leftarrow v_j \; if \, (j,i) \notin \pi; \; else \; if \, V_j \; is \; A \, , a'; else \; v_j^{\pi} \end{cases} \tag{5}$$

while keep everything else constant.

The $V_i^{\pi(a')}$(in short, V_i^{π}) is referred as the effect variable of this path intervention. From the defintion, we can see that comparing to *do*(or *info*) intervention in Def. 4, *path* intervention are considering both factual and couterfactual world for the cross-world nature of path-specific effects. Moreover, note that only the causal mechanisms on descendants of A could may change in the counterfactual world, thus

[1] i.i.d is short for independent and identically distributed.

effect variable $A^{\pi(a')} = A$ and $V_i^{\pi(a')} = V_i, j \notin \pi$. In fact, the intuition behind the path intervention is very simple, we are considering what effect will the information of $A = a'$ have (on the outcome Y) pass (only) through causal path π while keeping everything else constant, leading to a change of mechanism on the variables along the causal path. Comparing to the path-specific interventions in Def. 6 of [24], in which the π-specific counterfactual is simply assigning the treatment $do(a')$ exclusively to the causal path π while allowing all other causal paths to behave naturally, our path intervention counterfactual variable $Y^{\pi(a')}$ is simply making the treatment information $A = a'$ exclusively pass through the causal path π while allowing all other information outside from the π to remain the same as origin.

We use use the NIE(natural indirect effect) as an running example to show how our path intervention related to existing literatures.

Example 1 (Natural indirect effect) *For an SCM with three endogenous variables, including treatment A, mediator M, Outcome Y, satisfies*

$$\begin{cases} a \leftarrow f_A(u_A) \\ m \leftarrow f_M(a, u_M) \\ y \leftarrow f_Y(a, m, u_Y) \end{cases} \tag{6}$$

The natural indirect effect of Y on A through the mediator M can be defined as $Y(a, M(a'))$.

For a causal path $\pi : A \to M \to Y$ with transferred information $A = a'$, the path intervention results in a set of modified equations

$$\begin{cases} a \leftarrow f_A(u_A) \\ m \leftarrow f_M(a, u_M) \\ y \leftarrow f_Y(a, m, u_Y) \\ a^\pi \leftarrow f_A(u'_A) \\ m^\pi \leftarrow f_M(e'_{AM}, u'_M); e'_{AM} \leftarrow a' \\ y^\pi \leftarrow f_Y(e'_{AY}, e'_{MY}, u'_Y); e'_{AY} \leftarrow a, e'_{MY} \leftarrow m^\pi \end{cases} \tag{7}$$

Since the causal mechanisms of a and a^π coincide with the same input distribution information, the distributions of corresponding variables are the same. Henceforth, we marginalize A^π out the model, results in a causal model of variables of (A, M, Y, M^π, Y^π) with mechanisms:

$$\begin{cases} a \leftarrow f_A(u_A) \\ m \leftarrow f_M(a, u_M) \\ y \leftarrow f_Y(a, m, u_Y) \\ m^\pi \leftarrow f_M(e'_{AM}, u'_M); e'_{AM} \leftarrow a' \\ y^\pi \leftarrow f_Y(e'_{AY}, e'_{MY}, u'_Y); e'_{AY} \leftarrow a, e'_{MY} \leftarrow m^\pi \end{cases} \tag{8}$$

We can derive a set of structural equations ruling out $e'_{\cdot\cdot}$:

$$\begin{cases} a \leftarrow f_A(u_A) \\ m \leftarrow f_M(a, u_M) \\ y \leftarrow f_Y(a, m, u_Y) \\ m^\pi \leftarrow f_M(a', u'_M) \\ y^\pi \leftarrow f_Y(a, m^\pi, u'_Y) \end{cases} \tag{9}$$

We can see that causal mechanism for Y^π looks the same with $Y(a, M(a'))$, but there are different from two aspects. First, NIE usually consider the case of binary treatment value for two given value a, a' then mathematically $Y(a, M(a')) \neq Y(A, M(a')) = Y^{\pi(a')}$. Second, the interpretation is completely different for Y^π which is the effect of Y had the information $A = a'$ passed through the path $\pi : A \to M \to Y$ while keeping everything else unchanged. The presented distinctions in the above example suggests that our path-specific counterfactuals would be different from traditional ones which usually could be seen as generalizations of NIE. To further illustrate how the path intervention works, we introduce the intervention diagram as a graphical tool.

4. Intervention Diagrams

The recognition that there are mechanisms underlying the phenomena of interest, but that we usually cannot determine them precisely, gives rise to the discipline of causal inference [16]. Virtually every approach to causal inference works under the stringent condition that only partial knowledge of the underlying SCM is available, and usually the information of structural dependency(i.e. causal diagrams) might be sufficient for many causal effect estimation problems. Particularly, *do*-calculus is available for identification of causal queries when the causal structure is known, while its graphical criteria are placed on graphs which attained by removing input/output edges on intervened variables from the original causal graph. Since every hypothetical intervention is creating a counterfactual world for domain variables, then we can define a causal graph corresponding to the modified mechanisms. We here introduce a novel general concept of *intervention diagram* for interventions.

Definition 6 (Intervention diagram) *For an intervention, which can be* do*,* info *or* path *intervention, the causal diagram for the modified mechanisms induced by the intervention is called the* intervention diagram (or graph).

For some cases, we may also want to include exogenous variables to represent the causal mechanisms more detailed graphically, we call the causal graph of the intervention model with both exogenous and endogenous variables the *augmented intervention diagram*.

To emphasis the information account of causality and difference from previous causal graphs, our intervention diagrams, which is created based on the modified mechanisms by an intervention on SCM, will use lowercase letter to represent variables. To review all above concepts, the following example is used to present the intervention graphs for different interventions.

Example 2 *An SCM $\mathcal{M}$ with endogenous variables treatment T, outcome Y, confounder Z, two exogenous variables U_T, U_Z and an element t' in the domain of T, and its causal mechanisms represented by structural equations(the corresponding augmented causal graph is Fig 2(a)):*

$$\begin{cases} z \leftarrow f_Z(u_Z), \\ t \leftarrow f_T(z, u_T), \\ y \leftarrow f_Y(t, z). \end{cases}$$

Its do-intervention SCM $\mathcal{M}^{do(t')}$ with augmented intervention diagram Fig 2(b) has the following modified structural equations:

$$\begin{cases} z \leftarrow f_Z(u_Z), \\ t \leftarrow t', \\ y \leftarrow f_Y(t, z). \end{cases}$$

We reformulate $\mathcal{M}$ with an information interpretation:

$$\begin{cases} z \leftarrow f_Z(u_Z), \\ t \leftarrow f_T(e_{ZT}, u_T); e_{ZT} \leftarrow z, \\ y \leftarrow f_Y(e_{TY}, e_{ZY}); e_{TY} \leftarrow t, e_{ZY} \leftarrow z. \end{cases}$$

Its info-intervention SCM $\mathcal{M}^{\sigma(t')}$ with the augmented intervention diagram Fig 2(c) has the following modified structural equations:

$$\begin{cases} z \leftarrow f_Z(u_Z), \\ t \leftarrow f_T(e_{ZT}, u_T); e_{ZT} \leftarrow z, \\ y \leftarrow f_Y(e_{TY}, e_{ZY}); e_{TY} \leftarrow t', e_{ZY} \leftarrow z. \end{cases}$$

Its path-intervention SCM $\mathcal{M}^{\pi(t')}$(for causal path $\pi : T \rightarrow Y$ with input information $T = t'$) with augmented intervention diagram Fig 2(d) [2] has the following modified structural equations:

$$\begin{cases} z \leftarrow f_Z(u_Z), \\ t \leftarrow f_T(e_{ZT}, u_T); e_{ZT} \leftarrow z, \\ y \leftarrow f_Y(e_{TY}, e_{ZY}); e_{TY} \leftarrow t, e_{ZY} \leftarrow z, \\ y^\pi \leftarrow f_Y(e'_{TY}, e'_{ZY}); e'_{TY} \leftarrow t', e'_{ZY} \leftarrow z. \end{cases}$$

[2]Since t^π, z^π are not affected by the information $T = t'$ through the path π, then we marginalize out the mechanisms for them.

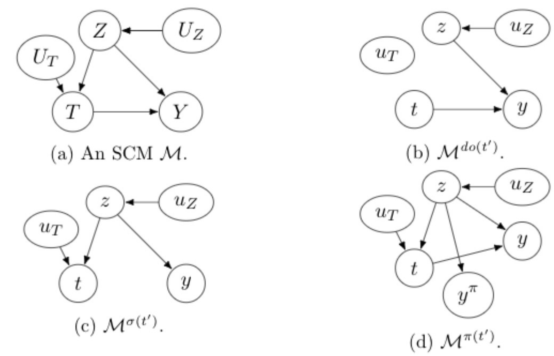

Figure 2. Causal graphs of an SCM and three augmented intervention diagrams of its intervention SCMs.

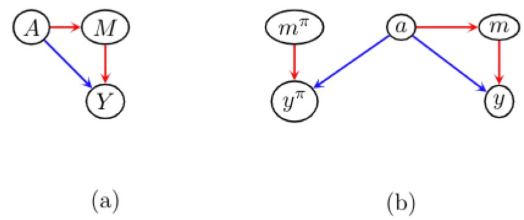

Figure 3. (a) Causal graph of treatment A, mediator M and outcome Y; (b) Intervention graph induced by $\pi(a')$.

Note that we usually construct the (augmented) causal graph of those intervention SCMs without separating the variables in the counterfactual world created by the hypothetical intervention from the factual world for conciseness, except for path intervention (which will be illustrated in the next example). In other worlds, the counterfactual variable $Y^{do(t')}$ created by do intervention $do(t')$ is represented with notation y instead of $y^{do(t')}$. In fact, from the definition of *do* intervention and *info* intervention, we are using y rather than $y^{do(t')}$ or $y^{\sigma(t')}$ to in those structural equations.

If we look into the definition of *do*, *info*, or *path* intervention for SCM in its information accounts, we will find that *path* intervention is quite different from *info* or *do* intervention in that it includes cross-world variables. In fact, cross-world quantities involved in mediation analysis is what makes path-specific effects complicated to deal with. The recanting witness criteria for identification of path-specific criteria is a proof for this complication. We now show that path interventions are usually need to add counterfactual variables along the path.

Back to Example 1(see Fig. 3), NIE could be deemed as the counterfactual variable Y^π for path intervention $\pi(a')$, where $\pi : A \rightarrow M \rightarrow Y$(red path in the Fig. 3(a)) with input information $A = a'$ instead of its factual value. According to the modified mechanisms Eq. 9, directed edges of $a \rightarrow m, m \rightarrow y, a \rightarrow y, a \rightarrow y^\pi, m^\pi \rightarrow y^\pi$ are in the intervention graph , see Fig. 3(b) for the intervention diagram of the $\pi(a')$ intervention SCM. We can see along the causal path $\pi : A \rightarrow M \rightarrow Y$, the path intervention has created a set of counterfactual variable M^π, Y^π. Someone might be confused with two things. First, why we are not using uppercase letter to denote variables in intervention diagrams? This is because we want to emphasis our interventions explicitly modify the assignment mechanisms of both factual and counterfactual variables. Second, for the intervention diagram induced by a path intervention, the counterfactual variables directly accept information from factual variables which can be read from the corresponding assignment mechanisms. This is different from *do* or *info* intervention in which all variables can be viewed as counterfactual variables (though some of them are not affected by the intervention, which have the same distribution with their corresponding factual variables). In other words, we usually have to place fact and counterfactual variable in the same graph which is similar to twin-SCM [17] [3].

Comparing to the interventionist approach of [20] to decompose treatment into separable components with nicotine-free cigarettes as example, our proposed path intervention explicitly describes the manipulation in structural causal model(SCM) with simple information transfer interpretation. Our counterfactual path-specific variable $Y^{\pi(a')}$ is interpreted as the effect of intervening the nicotine information $A = a'$ exclusively transferred on the causal path from treatment A to hypertensive status M and to Y while keep everything else constant. Moreover, our path-specific effects can even be well-defined for cyclic causal models while complications occur for recursive substitution approach in [20]. For example, an path intervention $\pi(a')$ on an SCM with cyclic causal graph Fig. 1 where $\pi : A \rightarrow M \rightarrow Y$, then the modified causal mechanisms are:

$$\begin{cases} a \leftarrow f_A(u_A) \\ m \leftarrow f_M(a, y, u_M) \\ y \leftarrow f_Y(m, u_Y) \\ m^\pi \leftarrow f_M(a', y, u'_M) \\ y^\pi \leftarrow f_Y(m^\pi, u'_Y) \end{cases} \tag{10}$$

Although our path-specific effects can be define for non-recursive SCMs, but we will not further address that problem in this paper considering the numerous complications for cyclic causal models. Thus in the rest of the paper, we only consider the case of Markovian acyclic SCMs i.e., non-parametric structural equation models with independent error terms(NPSEM-IE) [22].

[3] Our intervention diagram is apparently different from twin-SCM in the way of connecting factual and counterfactual variables.

Now we have defined the path interventions and the associated intervention diagrams for both counterfactual responses and factual variables, then how to identify effects of a path intervention? In particularly, how to identify path-specific effect for a path intervention $\pi(a')$ given the causal diagram but without knowledge of specific form of causal mechanisms $\{f_i\}_{i \in V}$ of an SCM $\mathcal{M}$? This is the main topic of our next section.

5. Identification of Path-specific Effects

Identification for traditional path-specific effects in a DAG is governed by a simple criterion known as the recanting witness criterion [1, 23, 22]. In contrast, our novel path-specific effects defined by our path intervention can be identified regardless of whether it satisfies recanting witness criteria or not. In particular, for a path intervention $\pi(a')$, the joint distribution of counterfactual variables can always be identified. To formulate our result, we denote $desc^\pi(A)$ as descendants of A in the path π, i.e., all nodes in π with a parent node in causal path π.

Theorem 7 (π-formula) *In a DAG $\mathcal{G}$(as the causal diagram of an SCM $\mathcal{M}$) with a factorized probability p, for a causal path $\pi : A \rightarrow \cdots \rightarrow Y$, the joint counterfactual distribution over variables $p(v^\pi, v)$ for path intervention $\pi(A = a')$ is identified via a equation called the π-formula:*

$$p(v^\pi_{desc^\pi(A)}, v) = \prod_{k \in desc^\pi(A)} p(v^\pi_k | e^\pi_{pa(k),k}) \cdot \prod_{i \in V} p(v_i | v_{pa(i)}) \tag{11}$$

where $e^\pi_{j,k} = v_j$ if $(j,k) \notin \pi$; else if V_j is A , a'; else v^π_j.

Proof By definition of path intervention, the modified causal mechanisms for $(\mathbf{V}^\pi, \mathbf{V})$ are, for any $i \in V$:

$$\begin{cases} v_i \leftarrow f_i(e_{pa(i),i}, u_i) \\ e_{j,i} \leftarrow v_j \\ v^\pi_i \leftarrow f_i(e'_{pa(i),i}, u'_i) \\ e'_{j,i} \leftarrow v_j \text{ if } (j,i) \notin \pi; \text{ else if } V_j \text{ is } A\ , a'; \text{else } v^\pi_j \end{cases}$$

The intervened SCM is also acyclic with independent errors, thus factorize with the following density:

$$p(v^\pi, v) = \prod_{k \in V} p(v^\pi_k | e^\pi_{pa(k),k}) \cdot \prod_{i \in V} p(v_i | v_{pa(i)})$$

where $e^\pi_{j,k} = v_j$ if $(j,k) \notin \pi$; else if V_j is A , a'; else v^π_j. Notice that for any counterfactual variables not in $desc^\pi(A)$, it receives the information directly from factual variables and the i.i.d copy of exogenous variable. These counterfactual variables are leafs of the intervention diagram, thus marginalization out them directly leads to the

π-formula. ∎

We now return to our running Example 1 along the path $A \to M \to Y$, the path-specific effects based on nested counterfactual can be identified. But when we add a variable Z with causal relation $Z \to A$ to the causal model results in the causal diagram in Fig 4(a), the path-specific effect along the path $Z \to A \to M \to Y$ can not be identified for the recanting witness A [4], and this seems to be awkward when path-specific effects of both path $Z \to A$ and path $A \to M \to Y$ can be identified. However, if we consider the path-specific effects defined by our path intervention, it can be identified with the π-formula in the following:

Example 3 *For a DAG for variables Z, A, M, Y, with causal relationships $Z \to A \to M \to Y$ and $A \to Y$ in Fig 4(a), the joint distribution of counterfactual and factual variables in 4(b) induced by path intervention $\pi(z')$, where $\pi : Z \to A \to M \to Y$, can be identified the following equation:*

$$
\begin{aligned}
&p(y^\pi, m^\pi, a^\pi, y, m, a, z) \\
&= p(v^\pi, v) \\
&= \prod_{k \in desc^\pi(Z)} p(v_k^\pi | e^\pi_{pa(k),k}) \cdot \prod_{i \in V} p(v_i | v_{pa(i)}) \\
&= p(y^\pi | a, m^\pi) p(m^\pi | a^\pi) p(a^\pi | z') p(y | a, m) p(m | a) p(a | z) p(z)
\end{aligned}
$$

Then the density of effect variable Y^π can be obtained by summation over $\{m^\pi, a^\pi, y, m, a, z\}$:

$$
\begin{aligned}
&p(y^\pi) \\
&= \sum_{m^\pi, a^\pi, m, a, z} p(y^\pi, m^\pi, a^\pi, y, m, a, z) \\
&= \sum_{m^\pi, a^\pi, a, z} p(y^\pi | a, m^\pi) p(m^\pi | a^\pi) p(a^\pi | z') \cdot p(a | z) p(z)
\end{aligned}
$$

The fact that our path-specific effect $Y^{\pi(z')}$ is mathematically different from $Y(z, z', \pi)$(see e.g. [23]) leads to different identification conditions. In many situations, the causal graphical model, with probability measure defined by π-formula for variables $(\mathbf{V}^\pi, \mathbf{V})$, can latent projection onto a set of variables of interest for identification. Let's see another more complicated example of recanting witness with latent projection onto variables of interest.

Example 4 *For a DAG (Fig.5(a))with a probability p that factorize, a path intervention $\pi(a')$ for the causal path $\pi : A \to M \to Y$ with input information $A = a'$, then according to intervention graph Fig. 5(b) with π-formula, we have*

[4]The definition of recanting witness and associated identification results can be found in [1, 10, 23, 22]

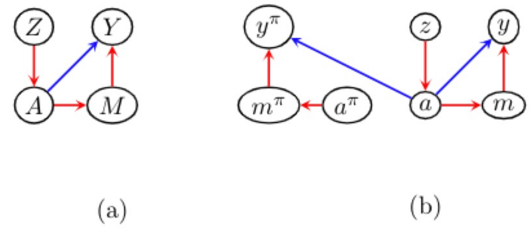

Figure 4. (a) Causal graph of Z, A, M, Y; (b) Intervention diagram induced by path intervention $\pi(z')$, where $\pi : Z \to A \to M \to Y$ with input information $Z = z'$.

$$
\begin{aligned}
&p(y^\pi, m^\pi, z, m, a, w) \\
&= p(y^\pi | m^\pi, w, z) p(y^\pi | w, a') \cdot p(z | m, a) p(m | a, w) p(a | w) p(w)
\end{aligned}
$$

Note that we use latent projection onto variables of interest, specifically, we exclude the factual variable Y in the intervention graph for simplicity which can be marginalized over the π-formula, but you cannot marginalize out factual variable M to identify the distribution of Y^π.

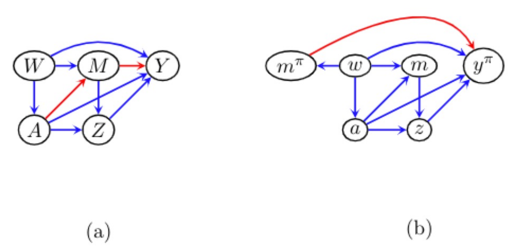

Figure 5. (a) A causal diagram with recanting witness M; (b) Intervention diagram of variables of interest.

This way of constructing intervention diagram is somewhat similar to twin-network [17] for both including factual and counterfactual variables. However, there are completely different in several ways. First, the twin-SCM only has U_V to connect factual and counterfactual variable, out intervention diagram includes an i.i.d copy of U_V for the corresponding intervened SCM. Second, there are no direct edges between factual and counterfactual variables in a twin-SCM. In contrast, there are no direct edges between two counterfacutal variables unless there is an edge on π connects them, which leads to counterfactual variables usually have no output edge in the intervention diagram. In other words, parents of counterfactual variables are usually factual variables, e.g. $A, Z, W \to Y^\pi$ in Fig. 5(b). Third, Our intervention diagrams include probabilistic relations instead of deterministic relations, while the incompleteness of d-separation occurs in twin-SCM due to deterministic relations [22].

6. Concluding remarks

By relating the informational account to SCMs, we propose a novel *path* intervention framework to formalize path-specific effects aligned the dictum "no causation without manipulation". In the *path* intervention framework, the information account of causality has suggested of separating causal mechanisms into information transfer and process. Our path intervention explicitly manipulate the mechanisms, thus path-specific effects even can be well-defined for cyclic SCMs. Intuitively, our path-specific effect $Y^{\pi(a')}$ is obtained by making the treatment information $A = a'$ exclusively pass through the causal path π while allowing all other information outside from the π to keep constant. Moreover, we address the identifiabilty of path-specific effects for acyclic Markovian SCMs and show that they can be identified even when recanting witness exists with the π-formula. In fact, it is obvious that our path intervention can easily be extended to a set of paths, and future work might be further explore the path-specific effects on cyclic SCMs.

References

[1] Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. 2005. 1, 6, 7

[2] E Bareinboim, JD Correa, D Ibeling, and T Icard. On pearl's hierarchy and the foundations of causal inference. *ACM Special Volume in Honor of Judea Pearl (provisional title)*, 2020. 2, 3

[3] Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986. 1

[4] Stephan Bongers, Jonas Peters, Bernhard Schölkopf, and Joris M Mooij. Theoretical aspects of cyclic structural causal models. *arXiv preprint arXiv:1611.06221*, 2016. 2

[5] John D Collier. Causation is the transfer of information. In *Causation and laws of nature*, pages 215–245. Springer, 1999. 1, 2, 3

[6] Bernard Feltz, Marcus Missal, and Andrew Sims. *Free Will, Causality, and Neuroscience*. Brill, 2019. 2

[7] Gong Heyang and Zhu Ke. Info intervention. *arXiv preprint arXiv:1907.11090*, 2019. 3

[8] Phyllis Illari and Federica Russo. *Causality: Philosophical theory meets scientific practice*. OUP Oxford, 2014. 2, 3

[9] Phyllis McKay Illari. Why theories of causality need production: An information transmission account. *Philosophy & Technology*, 24(2):95–114, 2011. 2

[10] Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright. *Handbook of graphical models*. CRC Press, 2018. 7

[11] Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. *Proceedings of machine learning research*, 89:3080, 2019. 1

[12] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995. 3

[13] Judea Pearl. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 2001. 1

[14] Judea Pearl. Causal and counterfactual inference. *The Handbook of Rationality*, pages 1–41, 2019. 2

[15] Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3):54–60, 2019. 1, 3

[16] Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 2000. 4

[17] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009. 2, 6, 7

[18] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018. 3

[19] James M Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, pages 143–155, 1992. 1

[20] James M Robins, Thomas S Richardson, and Ilya Shpitser. An interventionist approach to mediation analysis. *arXiv preprint arXiv:2008.06019*, 2020. 1, 3, 6

[21] Bernhard Schölkopf. Causality for machine learning. *arXiv preprint arXiv:1911.10500*, 2019. 2, 3

[22] Ilya Shpitser, Thomas S Richardson, and James M Robins. Multivariate counterfactual systems and causal graphical models. *arXiv preprint arXiv:2008.06017*, 2020. 6, 7, 8

[23] Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of statistics*, 44(6):2433, 2016. 1, 6, 7

[24] Junzhe Zhang and Elias Bareinboim. Non-parametric path analysis in structural causal models. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, 2018. 1, 4