Conditional Counterfactual Causal Effect for Individual Attribution

Ruiqi Zhao[1] Lei Zhang[2] Shengyu Zhu[3] Zitong Lu[4] Zhenhua Dong[3] Chaoliang Zhang[3] Jun Xu[2]
Zhi Geng[5] Yangbo He[*1]

[1]School of Mathematical Sciences, Peking University [2]Renmin University of China
[3]Huawei Noah's Ark Lab [4]City University of Hong Kong
[5]School of Mathematics and Statistics, Beijing Technology and Business University

Abstract

Identifying causes of an event, also termed as causal attribution, is a commonly encountered task in many applications. Existing methods, mostly in the Bayesian or causal inference literature, suffer from two main drawbacks: 1) they cannot attribute for individuals, and 2) only attribute one single cause at a time and cannot deal with the interaction effect among multiple causes. In this paper, we propose an individual causal attribution method, called conditional counterfactual causal effect (CCCE), which utilizes the individual observation as the evidence and is able to handle the common influence and interaction effects of multiple causes simultaneously. We discuss the identifiability of CCCE and then provide the identification formulas under certain proper assumptions. Finally, we conduct experiments on simulated and real data to illustrate the effectiveness of CCCE, compared with many state-of-the-art attribution methods.

1 INTRODUCTION

In many scenarios, we are interested in finding out the underlying causes of occurred events for individuals. For instance, in recommendation systems, when a user makes a purchase, the advertiser would like to know whether the purchase is caused by a particular advertising. This knowledge can be used to evaluate the effect of recommendation and further to guide advertisers to make reasonable payments and schedule future strategies. As another example, one may want to know the reasons why users give a low rating to a product, in order to develop better products. Addressing these questions, however, is not easy, as they require answering causal questions related to counterfactuals [Dawid, 2000]. Concretely, the advertisers want to know "Would the user have bought the item if there were no advertising?", while the latter is about "How would users have rated the product if some attributes of the product were different?". Therefore, we need a measure to quantify the change of the result when the cause changes, that is, how possibly the result is attributed to the cause.

For bivariate case (a cause and an effect), Pearl [2009] defined probability of necessity (PN) and probability of sufficiency (PS) to measure respectively how necessary or sufficient a cause is for the occurrence of the effect, and extended them to probability of necessary and sufficient causation (PNS), that is, how likely the event is affected in both ways. Dawid et al. [2014] defined probability of causation (PC) and related it to the causal relative risk (RR). These methods can only solve the attribution problem of two variables, but in practice, an effect can be affected by more than one causes. For the situation of multiple causes, Dawid et al. [2016] defined conditional probability of causation based on some background variables that are pre-treatment covariates. Jin [2012] discussed the identification of conditional causal effect. Lu et al. [2022] proposed the posterior total and direct causal effects (POSTTCE and POSTDCE) and extended the evidence (observational data of individuals) to the case of post-treatment variables. However, both conditional causal effect and POSTTCE can only measure the contribution of one cause to the outcome variable at a time, ignoring the joint influence or interaction effect of multiple causes. In practice, interaction effect can be important in many applications [Belch and Michael, 1998, Voorveld and Valkenburg, 2015, Jensen and Ruback, 1983, Ayres and Walter, 1991]. For example, in recommendation, the impact of using only television advertisement on users' purchase may be extremely limited, while using television and print advertisements at the same time may significantly improve the purchases [Naik and Raman, 2003, Lin et al., 2013]. In addition, the evidences used by conditional causal effect and POSTTCE are also limited. The former only focuses on the pre-treatment covariate, while the latter focuses on the outcome variable and the variables preceding the out-

[*]Correspondence to heyb@math.pku.edu.cn

Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023), PMLR 216:2519–2528.

come variable, omitting variables that may be affected by the outcome variable.

In this paper, for the case of multiple variables which may affect each other, we propose conditional counterfactual causal effect (CCCE). Different from PostTCE, CCCE can be used to measure the contribution of both a single cause and the combination of a set of causes simultaneously, and hence is able to characterize the interaction effect of two or more causes. Besides, the evidence on which CCCE is based can be any subset of all variables, including the pre-treatment variables, post-treatment variables, the outcome variable, and the variables that are affected by the outcome variable. We further discuss the relationship between CCCE and existing attribution measures. Notably, CCCE can be regarded as a generalization of the former, and also a generalization of PN and PC. Indeed, CCCE degenerates into PostTCE when considering a single cause on the outcome variable. Different from correlation and causal effect quantities, which can be identified by observational or interventional data, the measures of counterfactual inference are generally non-identifiable, even when randomized controlled experiments can be used. Therefore, we finally show the identifiability result and provide identification equations for CCCE under proper assumptions.

The rest of this paper is organized as follows. In Section 2, we present the notations used in this paper and review existing attribution methods. In Section 3, we propose CCCE and discuss their connections to existing methods. Section 4 discusses the idenfiability of CCCE and give the identification formulas. In Section 5, we use simulation experiments and real-data experiments to illustrate the effectiveness of CCCE. Finally, Section 6 concludes the paper.

2 PRELIMINARY

This section introduces useful notations, followed by a review of related attribution methods.

2.1 NOTATION

Unless otherwise stated, we use capital letters such as X to denote random variables, and use boldface letters like $\mathbf{X}$ to denote random variable sets or vectors. An instantiation of a variable or a vector is denoted by a lowercase letter, e.g., x and $\mathbf{x}$.

We consider binary variables $\{X_i\}_{i=1}^p =: \mathbf{X}$ as possible causes of an effect or outcome variable; here $X_i = 1$ for true (e.g., advertising) and $X_i = 0$ for false (e.g., non-advertising). Let a binary variable Y denote the outcome variable, and $Y = 1$ for true (e.g., purchase) and $Y = 0$ for false (e.g., no purchase). Note that X_i's are not necessarily independent and may affect each other. We use $\mathbf{Z} = \{Z_1, Z_2, \cdots, Z_q\}$ to denote variables which do not affect any variable of $\mathbf{X}$ and Y; similarly, $Z_i = 1$ for true (e.g., purchase of supporting equipment) and $Z_i = 0$ for false (e.g., no purchase of supporting equipment). We assume that variables $\mathbf{V} := \{\mathbf{X}, Y, \mathbf{Z}\}$ are given in a topological order, so that W_i is not affected by W_j for any $W_i, W_j \in \mathbf{V}$ with $i < j$. We can then divide $\mathbf{X}$ into two subsets according to the topological order w.r.t. $X_k \in \mathbf{X}$. That is, we can write $\mathbf{X} = \{\mathbf{A}_k, \mathbf{D}_k\}$, in which $\mathbf{A}_k = \{X_1, \cdots, X_{k-1}\}$ and $\mathbf{D}_k = \{X_k, \cdots, X_p\}$. Given a topological order, the data generating mechanism for $\mathbf{X}$ and Y can be described as

$$X_k = f_k(\mathbf{A}_k, \epsilon_k), k = 1, 2, \ldots, p, \text{ and } Y = f_Y(\mathbf{X}, \epsilon_Y),$$

where ϵ_k's and ϵ_Y are independent noise variables [Pearl, 2009]. The data generating mechanisms for variables in $\mathbf{Z}$ can be similarly defined.

For any subset $\mathbf{S} \subseteq \{1, \ldots, p\}$, denote $\mathbf{X}_\mathbf{S} = \{X_k : k \in \mathbf{S}\}$ and $|\mathbf{S}|$ as the cardinality of $\mathbf{S}$. Let $Y_{\mathbf{X}_\mathbf{S}=\mathbf{x}_\mathbf{S}}$ denote the potential outcome of Y under $\mathbf{X}_\mathbf{S} = \mathbf{x}_\mathbf{S}$, abbreviated as $Y_{\mathbf{x}_\mathbf{S}}$, and $(X_k)_{\mathbf{a}_k}$ is defined similarly. Let $\mathbf{X}_{-k} = \mathbf{X} \backslash \{X_k\}$ denote the set of variables X_i's without X_k, and $\mathbf{x} = \{x_1, \ldots, x_p\} \preceq \mathbf{x}' = \{x_1', \ldots, x_p'\}$ denote that $x_i \leq x_i'$ for all i. In this paper, we assume the *consistency* holds [Pearl, 2009], that is, for any variable sets $\mathbf{U}$ and $\mathbf{V}$, we have $\mathbf{U}_\mathbf{v} = \mathbf{U}$ if $\mathbf{V} = \mathbf{v}$ is observed. We also assume the *composition* holds [Pearl, 2009], that is, for any variable sets $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$, we have that $\mathbf{U}_\mathbf{v} = \mathbf{u}$ implies $\mathbf{W}_{\mathbf{uv}} = \mathbf{W}_\mathbf{v}$.

2.2 RELATED ATTRIBUTION METHODS

As discussed in [Dawid, 2000], assessing whether one event is the cause of another is different from measuring the effect of a cause. The latter commonly uses Bayesian methods or causal effect, which cannot answer attribution questions. Assume binary valued variables with 1 standing for true or that the event happens. The posterior probability approach is about association and cannot explain causation, since we can write $\mathrm{P}(X = 1 \mid Y = 1) = \mathrm{P}(Y = 1 \mid X = 1)\mathrm{P}(X = 1)/\mathrm{P}(Y = 1)$, which relies on the prior probability of X. The causal effect measures the effect of a cause, which is about the effect when a cause is turned on and is not suitable for finding the causes given an observed evidence. For example, poison can have a larger average causal effect than unhealthy food on the death. However, when we observe a person dies, it is not proper to always attribute the reason to poison, because he/she may not have taken poison at all. As such, finding the causes of an effect requires involving counterfactuals.

Consider cause X and effect Y, and $Y_{X=0}$ and $Y_{X=1}$ denote the potential outcomes of Y under $X = 0$ and $X = 1$, respectively. To measure how necessary X is a cause of an observed effect $Y = 1$, [Pearl, 2009] defined the probability of necessity as

$$\mathrm{PN}(X \Rightarrow Y) = \mathrm{P}(Y_{X=0} = 0 \mid X = 1, Y = 1), \qquad (1)$$

which gives the probability that the event Y would not have occurred in the absence of the event X given that both events Y and X did in fact occur. PN closely matches the reasoning used in lawsuits, where legal responsibility is understood counterfactually, that is, in the sense of necessary causation. In such a context, PN equals the probability that the damage Y suffered by the plaintiff would not have occurred were it not for the defendant's action X Pearl [1995]. The probability of sufficiency and the probability of necessity and sufficiency are defined as

$$\mathrm{PS}(X \Rightarrow Y) = \mathrm{P}(Y_{X=1} = 1 \mid X = 0, Y = 0), \qquad (2)$$
$$\mathrm{PNS}(X \Rightarrow Y) = \mathrm{P}(Y_{X=0} = 0, Y_{X=1} = 1). \qquad (3)$$

Pearl [2009] provided their identification equations under the assumption of monotonicity $Y_{X=0} \leq Y_{X=1}$ and the assumption of no confounding $X \perp Y_x$. Besides, Dawid et al. [2014] defined the probability of causation to measure how possible X is as a cause of an effect Y as

$$\mathrm{PC}(X \Rightarrow Y) = \mathrm{P}(Y_{X=0} = 0 \mid Y_{X=1} = 1). \qquad (4)$$

Note that $\mathrm{PC}(X \Rightarrow Y) = \mathrm{PN}(X \Rightarrow Y)$ if the assumption of no confounding holds. Therefore, PC and PN are simultaneously interventional and conditional probabilities. A limitation of PN, PS, PNS and PC is that they only consider the case of bivarate variables, that is, a cause and an effect. In practice, we can usually observe multiple other cause variables. Hence, Dawid et al. [2016] extended PC to the multivariate case, and defined the conditional probability of causation as $\mathrm{PC}(X \Rightarrow Y \mid \mathbf{W} = \mathbf{w}) = \mathrm{P}(Y_{X=0} = 0 \mid Y_{X=1} = 1, \mathbf{W} = \mathbf{w})$, in which W are pre-treatment covariates. Similarly, Lu et al. [2022] defined the conditional probability of necessity, and both of them can be identified under the assumption of monotonicity and the assumption of no confounding conditioned on W, that is, $X \perp Y_x \mid W$. In order to make full use of the observed evidence and solve the disadvantages of the causal effect methods, Lu et al. [2022] defined the posterior total causal effects based on the evidence about the known pre-treatment and post-treatment variables as follows

$$\begin{aligned} &\mathrm{PostTCE}(X_k \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = 1) \\ &= \mathrm{E}(Y_{X_k=1} - Y_{X_k=0} \mid \mathbf{X} = \mathbf{x}, Y = 1), \end{aligned} \qquad (5)$$

which measures the attribution of X_k on the effect for individuals of cases, that is, $Y = 1$, with the evidence $\mathbf{X} = \mathbf{x}$. Note that $\mathrm{PostTCE}(X_1 \Rightarrow Y \mid X_1 = x_1, Y = 1) = \mathrm{PN}(X_1 \Rightarrow Y)$ if there is only one cause, that is, $\mathbf{X} = \{X_1\}$. Hence, PostTCE can be considered as a generalization of PN in the case of multiple variables.

3 DEFINITION OF CONDITIONAL COUNTERFACTUAL CAUSAL EFFECT

Notice that PostTCE measures the impact of only *one* cause on the outcome. When there exist common effects or interaction effects between two or more causes, it fails to attribute multiple causes simultaneously. Besides, in the problem of individual attribution, we often take some known observational information of individuals as the evidence. PostTCE only takes the variables preceding the outcome as the evidence, and cannot use the information provided by the variables succeeding the outcome. To deal with these problems, we propose the conditional counterfactual causal effect, which measures the causal effect of $\mathbf{X_S}$ on Y for individuals with the evidence $\mathbf{W} = \mathbf{w}$.

Definition 1 (Conditional Counterfactual Causal Effect)**.** *Given the variable set $\mathbf{V} = (\mathbf{X}, Y, \mathbf{Z})$, the evidence $\mathbf{W} = \mathbf{w}$ and a set of causes $\mathbf{X_S} \subseteq \mathbf{X}$, the conditional counterfactual causal effect of $\mathbf{X_S}$ on Y is defined as*

$$\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{W} = \mathbf{w}) = \mathbb{E}\left(Y_{\mathbf{x}^1_{\mathbf{S}}} - Y_{\mathbf{x}^0_{\mathbf{S}}} \mid \mathbf{W} = \mathbf{w}\right), \qquad (6)$$

in which $\mathbf{W} \subseteq \mathbf{V}$ and $\mathbf{x}^1_{\mathbf{S}} \succeq \mathbf{x}^0_{\mathbf{S}}$.

In many applications such as those mentioned in Section 1, we would set $\mathbf{x}^1_{\mathbf{S}}$ and $\mathbf{x}^0_{\mathbf{S}}$ as $\mathbf{1}_{|\mathbf{S}|}$ and $\mathbf{0}_{|\mathbf{S}|}$, which are vectors with all entries being 1 and 0, respectively. By this definition, CCCE can measure the joint influence of all causes in $\mathbf{X_S}$ on the outcome Y. We remark that $\mathbf{x}^1_{\mathbf{S}}$ and $\mathbf{x}^0_{\mathbf{S}}$ can be set to other values given other cases of interest; in this paper we will focus on the case with $\mathbf{x}^1_{\mathbf{S}}$ and $\mathbf{x}^0_{\mathbf{S}}$ being $\mathbf{1}_{|\mathbf{S}|}$ and $\mathbf{0}_{|\mathbf{S}|}$, respectively. In addition, the evidence $\mathbf{W} = \mathbf{w}$ always contains the observation of Y. For example, in order to evaluate the effect of advertising, we are more concerned about the conversion rate of recommendations among people with buying behavior ($Y = 1$). Hence, CCCE measures the causation for causes in $\mathbf{X_S}$. The larger the CCCE of $\mathbf{X_S}$ on Y is, the larger the attribution of the effect to the causes $\mathbf{X_S}$ is. Note that the evidence $\mathbf{W} = \mathbf{w}$ can contain the observation of $\mathbf{X_S}$. Therefore, CCCE is different from the conditional causal effect in Jin [2012]. The former is a counterfactual variable, while the latter only conditions on covariates except $\mathbf{X_S}$, that is, the conditional causal effect only involves intervention variables and can be identified with observational and interventional data under suitable assumptions.

CCCE can be regarded as a generalization of PostTCE. Specifically, in the case of $\mathbf{Z} = \varnothing$ and $\mathbf{X_S} = \{X_1\}$, if the evidence $\mathbf{W} = \mathbf{w}$ is in the form of $\{\mathbf{U} = \mathbf{u}, Y = 1 : \mathbf{U} \subseteq \mathbf{X}\}$, then CCCE is reduced to $\mathrm{E}(Y_{X_1=1} - Y_{X_1=0} \mid \mathbf{U} = \mathbf{u}, Y = 1)$, which is exactly $\mathrm{PostTCE}(X_1 \Rightarrow Y \mid \mathbf{U} = \mathbf{u}, Y = 1)$ [Lu et al., 2022]. The details will be given in next section. Further, in the case of $\mathbf{Z} = \varnothing$ and $\mathbf{X_S} = \{X_1\}$, if the evidence is given by $\{X_1 = 1, Y = 1\}$, then CCCE is reduced to $\mathrm{E}(Y_{X_1=1}-Y_{X_1=0} \mid X_1 = 1, Y = 1) = \mathrm{P}(Y_{X_1=0} = 0 \mid X_1 = 1, Y = 1)$, which is exactly $\mathrm{PN}(X_1 \Rightarrow Y)$ [Pearl, 2009]. What's more, if the evidence is given by $\{X_1 = 0, Y = 0\}$, then CCCE is reduced to $\mathrm{E}(Y_{X_1=1} - Y_{X_1=0} \mid X_1 = 0, Y = 0) = \mathrm{P}(Y_{X_1=1} = 1 \mid X_1 = 0, Y = 0)$, which is exactly $\mathrm{PS}(X_1 \Rightarrow Y)$ [Pearl, 2009]. Thus, CCCE can be considered as a generalization of PN, PS and POSTTCE. It should be noted that this generalization is not trivial, because CCCE calculates the impact of random multiple causes on the result simultaneously, and also adds $\mathbf{Z}$ into the evidence set, which makes the identifiability of CCCE more complex than POSTTCE.

4 IDENTIFIABILITY FOR CONDITIONAL COUNTERFACTUAL CAUSAL EFFECT

Similar to PN, PC and POSTTCE, CCCE also requires to calculate the probability of counterfactual variables. In this section, we provide the assumptions required for identifiability (Section 4.1) and discuss the identifiability of CCCE under different evidences. In Section 4.2, we present the identification formula of CCCE given the evidence about a subset of $\mathbf{X}$. In Section 4.3, we discuss the idenfiability of CCCE given the evidence about the outcome variable Y and a subset of $\mathbf{X}$. In Section 4.4, we extend the evidence set to the general case, that is, an arbitrary subset $\mathbf{W} \subseteq \mathbf{V}$, and give a general result on the identifiability of CCCE.

4.1 ASSUMPTIONS

To give the identification formula of CCCE, the following assumptions are required.

Assumption 1 (No confounding)**.**

(a) There is no confounding among the variables in $\mathbf{X}$, that is, $(\mathbf{X}_k)_{\mathbf{a}_k} \perp \mathbf{A}_k$ for all $\mathbf{a}_k$ and $k = 2, \ldots, p$;

(b) There is no confounding between Y and $\mathbf{X}$, that is, $Y_{\mathbf{x}} \perp \mathbf{X}$ for all $\mathbf{x}$;

(c) Given $\mathbf{X}$ and Y, there is no confounding between $(\mathbf{X}, Y)$ and $\mathbf{Z}$, that is, $(Y_{\mathbf{x}}, \mathbf{X}_{\mathbf{x_S}}) \perp \mathbf{Z} \mid \mathbf{X}, Y$ for all $\mathbf{x}, \mathbf{x_S}$ and $\mathbf{X_S} \subset \mathbf{X}$.

The assumption of no confounding is also known as the assumption of ignorability [Paul and Donald, 1983] or exogeneity [Pearl, 2009], implying that there is no unobserved confounders. The Assumptions 1(a) and 1(b) are satisfied if $\epsilon_1, \ldots, \epsilon_p$ are mutually independent and $(\epsilon_1, \ldots, \epsilon_p)$ and ϵ_Y are independent, respectively, while the Assumption 1(c) is satisfied when $(\epsilon_1, \ldots, \epsilon_p, \epsilon_Y)$ and the noise variables of $\mathbf{Z}$ are independent.

Assumption 2 (Monotonicity)**.**

(a) The variables in $\mathbf{X}$ satisfy the monotonicity, that is, $(X_k)_{\mathbf{a}_k} \leq (X_k)_{\mathbf{a}'_k}$ for all $k = 1, \ldots, p$ whenever $\mathbf{a}_k \preceq \mathbf{a}'_k$;

(b) The outcome variable Y satisfies the monotonicity, that is, $Y_{\mathbf{x}} \leq Y_{\mathbf{x}'}$ whenever $\mathbf{x} \preceq \mathbf{x}'$.

The assumption of monotonicity is often assumed in practice, implying that the causes cannot prevent the effect. For example, in the field of recommendation, the effect of using television and print advertisements at the same time is no worse than using only one of them. In this sense, the effect of $n+1$ causes is definitely not worse than n causes. However, it should be noted that there are many redundant causes. For example, given all parents of the outcome, the other ancestor variables become independent of the outcome. In reality, there indeed exists some situations that the monotonicity cannot be satisfied. For example, moderate exercise is beneficial for physical health, but excessive exercise can actually be harmful. But in general, monotonicity is a common assumption in attribution problems and can be satisfied in most cases in practical applications.

4.2 IDENTIFIABILITY OF THE CCCE CONDITIONED ON A SUBSET OF X

Under Assumptions 1 and 2, we discuss the identifiability and provide the identification equation of CCCE given the evidence about a subset of $\mathbf{X}$, that is, $\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X}' = \mathbf{x}')$, where $\mathbf{X}' \subseteq \mathbf{X}$. According to the definition of CCCE, we first discuss the identifiability of the conditional probabilities $\mathrm{P}(Y_{\mathbf{x}^1_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x})$ and $\mathrm{P}(Y_{\mathbf{x}^0_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x})$ for any $\mathbf{x}^1_\mathbf{S} \succeq \mathbf{x_S} \succeq \mathbf{x}^0_\mathbf{S}$, where $\mathbf{x_S} \subseteq \mathbf{x}$, as shown in Lemma 1 and Lemma 2 in the following.

Lemma 1. *Given a causal ordering $(\mathbf{X}, Y, \mathbf{Z}) = (X_1, \ldots, X_p, Y, \mathbf{Z})$, let $\mathbf{S} \subseteq \{1, \ldots, p\}$ and $\mathbf{x}^0_\mathbf{S} \preceq \mathbf{x_S}$. Under Assumptions 1(a), 1(b) and Assumption 2(a), the conditional probability $P(Y_{\mathbf{x}^0_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x})$ is identifiable, and its identification formula is*

$$
\begin{aligned}
&\mathrm{P}(Y_{\mathbf{x}^0_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x}) \\
=&\mathrm{P}(Y_{\mathbf{x}^0_\mathbf{S}} = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{D}_k = \mathbf{d}_k) \\
=& \sum_{\mathbf{c}_{k:p} \preceq \mathbf{d}_k} \Bigg\{ \mathrm{P}(Y = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{D}_k = \mathbf{c}_{k:p}) \\
&\times \prod_{i \in \{k, \ldots, p\} \setminus \mathbf{S}} \Bigg[1 - x_i c_i + x_i (-1)^{1-c_i} \\
&\times \frac{\mathrm{P}(X_i = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{X}_{k:i-1} = \mathbf{c}_{k:i-1})}{\mathrm{P}(X_i = x_i \mid \mathbf{A}_i = \mathbf{a}_i)} \Bigg] \Bigg] \Bigg\},
\end{aligned}
\tag{7}
$$

where $\mathbf{x_s} \subseteq \mathbf{x}$, $k = \min \mathbf{S}$, $\mathbf{X}_{k:i-1} = \{X_k, \ldots, X_{i-1}\}$ and $\mathbf{c}_{k:p} = (c_k, \ldots, c_p)$ satisfying $c_i = x^0_i$ if $i \in \mathbf{S}$.

Proof (sketch). By topological order, starting from the variable with the smallest index in $\mathbf{X_S}$, iteratively simplify the counterfactual terms into the observational terms by using the no confounding assumption, monotonicity assumption, consistency and composition. More details are given in the supplementary material. □

Lemma 2. *Given a causal ordering* $(\mathbf{X}, Y, \mathbf{Z}) = (X_1, \dots, X_p, Y, \mathbf{Z})$*, let* $\mathbf{S} \subseteq \{1, \dots, p\}$ *and* $\mathbf{x}^1_{\mathbf{S}} \succeq \mathbf{x}_{\mathbf{S}}$*. Under Assumptions 1(a), 1(b) and Assumption 2(a), the conditional probability* $P(Y_{\mathbf{x}^1_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x})$ *is identifiable, and its identification formula is*

$$\begin{aligned}
&\mathrm{P}(Y_{\mathbf{x}^1_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x}) \\
=&\mathrm{P}(Y_{\mathbf{x}^1_{\mathbf{S}}} = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{D}_k = \mathbf{d}_k) \\
=&\sum_{\mathbf{c}_{k:p} \succeq \mathbf{d}_k} \Bigg\{ \mathrm{P}(Y = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{D}_k = \mathbf{c}_{k:p}) \\
&\times \prod_{i \in \{k, \dots, p\} \backslash \mathbf{S}} \Bigg[x_i + c_i - x_i c_i + (1 - x_i)(-1)^{c_i} \\
&\times \frac{\mathrm{P}(X_i = 0 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{X}_{k:i-1} = \mathbf{c}_{k:i-1})}{\mathrm{P}(X_i = x_i \mid \mathbf{A}_i = \mathbf{a}_i)} \Bigg] \Bigg\},
\end{aligned} \tag{8}$$

where $\mathbf{x_s} \subseteq \mathbf{x}, k = \min \mathbf{S}, \mathbf{X}_{k:i-1} = \{X_k, \dots, X_{i-1}\}$ *and* $\mathbf{c}_{k:p} = (c_k, \dots, c_p)$ *satisfying* $c_i = x^1_i$ *if* $i \in \mathbf{S}$.

A proof (and also proofs of the remaining lemmas and theorems) can be found in the supplementary material. Taking Lemma 1 as an example, the conditional probability $\mathrm{P}(Y = 1 \mid \mathbf{A}_k = \mathbf{a}_k, \mathbf{D}_k = \mathbf{c}_{k:p})$ in Equation (7) contains $\mathbf{D}_k$ in the condition part, which may be affected by $\mathbf{X}_{\mathbf{S}}$. Especially, if the observed sample $\mathbf{x}_{\mathbf{S}} = \mathbf{x}^0_{\mathbf{S}}$ appears in the evidence $\mathbf{W} = \mathbf{w}$, then the conditional probability $\mathrm{P}(Y_{\mathbf{x}^0_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x}) = \mathrm{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ by consistency. If $\mathbf{Z} = \varnothing$ and $|\mathbf{S}| = 1$, that is, $\mathbf{X}_{\mathbf{S}} = \{X_1\}$, then Lemma 1 degenerates to Lemma 1 in Lu et al. [2022]. A similar observation holds for Lemma 2.

Based on the above lemmas, we next present the identifiability result of CCCE given the evidence about $\mathbf{X}$.

Theorem 1. *Given a causal ordering* $(X_1, \dots, X_p, Y, \mathbf{Z})$*, let* $\mathbf{S} \subseteq \{1, \dots, p\}$*. Under Assumptions 1(a), 1(b) and Assumption 2(a),* $\mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X} = \mathbf{x})$ *is identifiable, and its identification formula can be obtained by substituting the equations in Lemma 1 and Lemma 2 into its definition, that is,*

$$\begin{aligned}
&\mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}) \\
=&\mathrm{P}(Y_{\mathbf{x}^1_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x}) - \mathrm{P}(Y_{\mathbf{x}^0_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x}),
\end{aligned} \tag{9}$$

where $\mathbf{x}^1_{\mathbf{S}} \succeq \mathbf{x}_{\mathbf{S}} \succeq \mathbf{x}^0_{\mathbf{S}}$ *and* $\mathbf{x}_{\mathbf{S}} \subseteq \mathbf{x}$.

Notice that the identification formulas in Lemma 1 and Lemma 2 only have the probability of observed variables, implying that CCCE in Theorem 1 can be estimated from the observational data only. However, sometimes we can only get a part of evidence of $\mathbf{X}$, say, $\mathbf{X}' \subseteq \mathbf{X}$. Hence, we show below the identifiability of the CCCE based on the evidence about a subset of $\mathbf{X}$.

Corollary 1. *Given a causal ordering* $(X_1, \dots, X_p, Y, \mathbf{Z})$*, let* $\mathbf{S} \subseteq \{1, \dots, p\}, \mathbf{X}' \subseteq \mathbf{X}$ *and* $\mathbf{X}' = \mathbf{x}'$*, then the following equation holds:*

$$\begin{aligned}
&\mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X}' = \mathbf{x}') \\
=&\sum_{\mathbf{x}:\mathbf{x} \supseteq \mathbf{x}'} \mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}) \\
&\times \mathrm{P}(\mathbf{X} = \mathbf{x} \mid \mathbf{X}' = \mathbf{x}').
\end{aligned} \tag{10}$$

Further, if Assumptions 1(a), 1(b) and 2(a) are satisfied, $\mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X}' = \mathbf{x}')$ *is identifiable according to Theorem 1.*

In the corollary above, we show that the CCCE with the evidence of a subset $\mathbf{X}' = \mathbf{x}'$ is the expectation of that with the evidence of the full set $\mathbf{X} = \mathbf{x}$ that are taken over the unknown causes $\mathbf{X} \backslash \mathbf{X}'$ conditionally on $\mathbf{X}' = \mathbf{x}'$. In other words, the CCCE with the evidence of a subset $\mathbf{X}' = \mathbf{x}'$ is identifiable whenever that with the evidence of the full set $\mathbf{X} = \mathbf{x}$ is identifiable.

4.3 IDENTIFIABILITY OF THE CCCE CONDITIONED ON $(\mathbf{X}, Y)$

In the practical application of the attribution problem, we often pay more attention to the group of cases ($Y = 1$). For example, among all the people who buy an item, advertisers want to know whether their buying behavior is due to recommendation. Below we give the identifiability of CCCE given the evidence about $\mathbf{X}$ and Y.

Theorem 2. *Under Assumption 2(b), the following equations hold:*

$$\begin{aligned}
&\mathrm{CCCE}(\mathbf{X}_S \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = 1) \\
=&1 - \frac{\mathrm{P}(Y_{\mathbf{x}^0_{\mathbf{S}}} = 1 \mid \mathbf{X} = \mathbf{x})}{\mathrm{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})};
\end{aligned} \tag{11}$$

$$\begin{aligned}
&\mathrm{CCCE}(\mathbf{X}_S \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = 0) \\
=&1 - \frac{\mathrm{P}(Y_{\mathbf{x}^1_{\mathbf{S}}} = 0 \mid \mathbf{X} = \mathbf{x})}{\mathrm{P}(Y = 0 \mid \mathbf{X} = \mathbf{x})},
\end{aligned} \tag{12}$$

where $\mathbf{S} \subseteq \{1, \dots, p\}, \mathbf{x}^1_{\mathbf{S}} \succeq \mathbf{x}_{\mathbf{S}} \succeq \mathbf{x}^0_{\mathbf{S}}$ *and* $\mathbf{x}_{\mathbf{S}} \subseteq \mathbf{x}$*. Further, if Assumptions 1(a), 1(b) and 2(a) are satisfied,* $\mathrm{CCCE}(\mathbf{X}_{\mathbf{S}} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = y)$ *for* $y = 0, 1$ *is identifiable and its identification formulas can be obtained by substituting the equations 7 and 8 into equations above, respectively.*

Similar to Corollary 1, we also present the identifiability of CCCE based on the evidence about the outcome variable Y and a subset of $\mathbf{X}$.

Corollary 2. *Given a causal ordering* $(X_1, \ldots, X_p, Y, \mathbf{Z})$, *let* $\mathbf{S} \subseteq \{1, \ldots, p\}, \mathbf{X}' \subseteq \mathbf{X}$ *and* $\mathbf{X}' = \mathbf{x}'$, *then the following equation holds:*

$$\begin{aligned}&\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X}' = \mathbf{x}', Y = y)\\ =&\sum_{\mathbf{x}:\mathbf{x}\supseteq\mathbf{x}'} \mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = y)\\ &\times \mathrm{P}(\mathbf{X} = \mathbf{x} \mid \mathbf{X}' = \mathbf{x}', Y = y),\end{aligned} \tag{13}$$

for $y = 0, 1$. *Further, if Assumptions 1(a), 1(b) and 2 are satisfied,* $\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X}' = \mathbf{x}', Y = y)$ *is identifiable according to Theorem 2.*

4.4 A GENERAL IDENTIFIABILITY RESULT FOR CCCE

Given the previous identifiability results, we now consider a more general setting and discuss the identifiability of CCCE given the evidence about an arbitrary subset. Before describing the main result in Theorem 3, we start with two useful results.

Lemma 3. *Under Assumption 1(c), we have*

$$\begin{aligned}&\mathrm{P}\left(Y_{\mathbf{x}^*_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x}, Y = y, \mathbf{Z} = \mathbf{z}\right)\\ =&\mathrm{P}\left(Y_{\mathbf{x}^*_\mathbf{S}} = 1 \mid \mathbf{X} = \mathbf{x}, Y = y\right),\end{aligned}$$

where $\mathbf{S} \subseteq \{1, \ldots, p\}$ *and* $\mathbf{x}^*_\mathbf{S}$ *denotes an arbitrary value of* $\mathbf{x_S}$.

Using Lemma 3, we can give the identifiability of CCCE based on the evidence about the complete set $(\mathbf{X}, Y, \mathbf{Z})$, as shown in the following.

Corollary 3. *Given a causal ordering* $(X_1, \ldots, X_p, Y, \mathbf{Z})$, *let* $\mathbf{S} \subseteq \{1, \ldots, p\}$, CCCE *based on the evidence about the complete set has the following equation:*

$$\begin{aligned}&\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = y, \mathbf{Z} = \mathbf{z})\\ =&\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = y).\end{aligned} \tag{14}$$

Based on all conclusions in this section, now we can give a general result for the identifiability of CCCE based on the evidence about an arbitrary subset of $\mathbf{V}$, as shown below.

Theorem 3. *Given a causal ordering* $(X_1, \ldots, X_p, Y, \mathbf{Z})$, *let* $\mathbf{S} \subseteq \{1, \ldots, p\}$, *and* $\mathbf{W}$ *is an arbitrary subset of* $(\mathbf{X}, Y, \mathbf{Z})$. *Under Assumption 1 and Assumption 2,* CCCE *of* $\mathbf{X_S}$ *on* Y *based on the evidence* $\mathbf{W} = \mathbf{w}$ *has the following equation:*

$$\begin{aligned}&\mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{W} = \mathbf{w})\\ =&\sum_{(\mathbf{x},y,\mathbf{z}):(\mathbf{x},y,\mathbf{z})\supseteq\mathbf{w}} \mathrm{CCCE}(\mathbf{X_S} \Rightarrow Y \mid \mathbf{X} = \mathbf{x}, Y = y)\\ &\times \mathrm{P}(\mathbf{X} = \mathbf{x}, Y = y, \mathbf{Z} = \mathbf{z} \mid \mathbf{W} = \mathbf{w}),\end{aligned} \tag{15}$$

which is identifiable according to Theorem 2.

According to the results in this section, CCCE only uses the topological ordering of the variables for the attribution, but a causal graph may has several different valid topological orderings. In fact, for a given graph, the value of CCCE is invariant for different valid topological orderings, as long as the evidence set $\mathbf{W} = \mathbf{w}$ contains the variable Y and its all ancestors. Note that, for any topological ordering of a given graph, the order of ancestors of Y always precedes the order of Y. Therefore, for any given valid topological ordering, we only need to make the evidence set $\mathbf{W} = \mathbf{w}$ contain Y and the variables before Y in this ordering.

5 EXPERIMENTS

In this section, we conduct experiments to illustrate the effectiveness of CCCE and compare it with other attribution methods. We run the experiments on a desktop computer with Intel(R) Core(TM) i5-8250U CPU and 8GB RAM. The codes are available at `https://github.com/LLily0703/CCCE`.

5.1 SIMULATION

Setup Consider a real world scenario in which we aim to find out why a family purchases eye-protection lamps. The causal graph describing the relations of related factors is given in Figure 1, where Children (X_1), Kid Tablets (X_2), Education (X_3), Books (X_4), Children's Watch (X_5) and Stationery (X_6) are some possible causes of Eye-protection Lamps (Y) and the direct arrows indicate direct effects between variables. All the variables are binary-valued, and $X_1 = 1$ denotes that there are children in the family, $X_3 = 1$ denotes the family attaches importance to education, $X_j = 1, j = 2, 4, 5, 6$ denotes that the family has purchased the corresponding items and $Y = 1$ denotes the family has purchased the eye-protection lamps. According to the causal graph, a topological order of variables is given by $(X_1, X_3, X_2, X_4, X_5, Y, X_6)$. Note that while the true causal graph is used for generating data in our experiments, only this topological order is available for attribution. Besides, we assume that these variables satisfy the assumption of no confounding and the assumption of monotonicity.

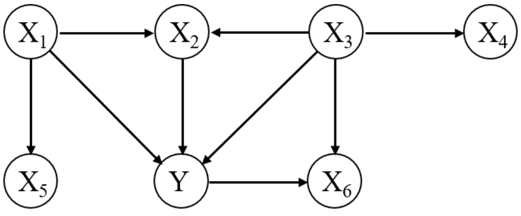

Figure 1: A causal graph of eye-protection lamps

Data Generating Process The do-calculus [Pearl, 1995] and the Markov property of $\mathrm{P}(\mathbf{V})$ w.r.t. causal graph $\mathcal{G}$

Table 1: Change rate of methods for attribution without interaction effect.

	RAND	POST	PN	PS	PNS	ACE	POSTTCE	CCCE
CR1	0.2390	0.4502	0.4598	0.4437	0.4602	0.4515	0.6636	**0.6713**
std.	(0.0143)	(0.0278)	(0.0247)	(0.0419)	(0.0276)	(0.0284)	(0.0290)	(0.0303)
CR2	0.3842	0.6706	0.6434	0.6420	0.6394	0.6712	0.7026	**0.7398**
std.	(0.0141)	(0.0360)	(0.0257)	(0.0268)	(0.0267)	(0.0350)	(0.0277)	(0.0200)

in Figure 1 allow the following factorization of the joint distribution [Lauritzen, 1996]:

$$\begin{aligned}
&\mathrm{P}(\mathbf{V}) \\
&= \mathrm{P}(X_1)\mathrm{P}(X_3)\mathrm{P}(X_2 \mid X_1, X_3)\mathrm{P}(X_4 \mid X_3)\mathrm{P}(X_5 \mid X_1) \\
&\quad \times \mathrm{P}(X_6 \mid X_3)\mathrm{P}(Y \mid X_1, X_2, X_3) \\
&= \mathrm{P}(X_1)\mathrm{P}(X_3)\mathrm{P}((X_2)_{X_1,X_3})\mathrm{P}((X_4)_{X_3})\mathrm{P}((X_5)_{X_1}) \\
&\quad \times \mathrm{P}((X_6)_{X_3})\mathrm{P}(Y_{X_1,X_2,X_3}).
\end{aligned}$$

Let $\mathrm{PA}(X, \mathcal{G})$ denote the parents of X in the causal graph $\mathcal{G}$, which can be abbreviated as $\mathrm{PA}(X)$. Then the conditional distribution of each variable X_i is transformed into the marginal distribution of $2^{|\mathrm{PA}(X_i,\mathcal{G})|}$ binary counterfactual variables, except for the root nodes. Hence, according to the topological order, we can generate observation data by using these counterfactual probabilities. First, we generate samples (x_1, x_3) of root variables (X_1, X_3) according to the distributions $\mathrm{P}(X_1), \mathrm{P}(X_3)$. Then, according to the topological ordering, generate the sample x_j of X_j by their marginal counterfactual distributions $\mathrm{P}((X_j)_{\mathrm{PA}(X_j)=\mathrm{pa}(X_j)})$ in order, where $\mathrm{pa}(X_j)$ is the sample of the parent set of X_j we have already generated. For example, we first generate the samples $(X_1, X_3) = (x_1, x_3) = (1, 0)$, then we can generate the sample x_2 of X_2 by its marginal counterfactual distribution $\mathrm{P}((X_2)_{x_1,x_3}) = \mathrm{P}((X_2)_{1,0})$. In fact, the sample we generate here is the sample $(x_2)_{x_1,x_3} = (x_2)_{1,0}$ of the counterfactual variable $(X_2)_{x_1,x_3} = (X_2)_{1,0}$, but according to the consistency assumption, when we observe $(X_1, X_3) = (x_1, x_3) = (1, 0)$, we have $(X_2)_{x_1,x_3} = (X_2)_{1,0} = X_2$. Hence, we can get $(x_2)_{x_1,x_3} = (x_2)_{1,0} = x_2$ here. By iterating this step, we can generate samples of all variables in order. In our simulation, the probabilities are either (i) randomly generated, or (ii) generated by logistic functions:

$$\begin{aligned}
&\mathrm{P}((X_j)_{\mathrm{pa}(X_j)} = 1) = \text{logistic}(\gamma_j + \alpha_j^T \cdot \mathrm{pa}(X_j)); \\
&\mathrm{P}(Y_{x_1,x_2,x_3} = 1) = \text{logistic}(\gamma_Y + \alpha_{Y1}x_1 + \alpha_{Y2}x_2 \\
&\qquad\qquad\qquad\qquad + \alpha_{Y3}x_3 + \beta x_2 x_3),
\end{aligned} \tag{16}$$

where $j = 1, \ldots, 6$, $\mathrm{pa}(X_j)$ denotes the value of $\mathrm{PA}(X_j)$, $\gamma_1, \cdots, \gamma_6, \gamma_Y$ are intercept terms, $\alpha_1, \cdots, \alpha_6, \alpha_Y$ are vectors of weights, and β characterizes the interaction effect between X_2 and X_3.

According to the above data generating process, we generate ten different causal graphs with the same structure but different causal mechanisms (edge weights). Afterwards, we conducte ten experiments on each causal graph, that is, generate ten different datasets with a sample size of 1000. We first randomly select the probabilities. For each individual sample, we use RAND (randomly select in variables preceding Y), POST ($\mathrm{P}(X_k = 1 \mid Y = 1)$), PN, PS, PNS, ACE, POSTTCE and CCCE for attribution and take the observational sample $(x_1, x_2, x_3, x_4, y, x_6)$ as the evidence. When looking for only one reason, we regard the variable with the largest value as the true cause of the outcome. For two causes, CCCE can measure the impact of two variables simultaneously, while other methods measure one cause at a time and find the two with the first two largest values as true causes.

Evaluation and Result After finding the causes, we use the Change Rate (CR, higher is better) to measure the effectiveness and accuracy of attribution. To define the Change Rate, let N be the size of observational data set $\mathcal{D}$. Let $\mathcal{A} = \{v = (x_1, x_3, x_2, x_4, x_5, y, x_6) \in \mathcal{D} \mid y = 1 \text{ occurs in } v\}$. Then for any vector in $\mathcal{A}$, set the value of cause variables $\mathbf{X_S}$ in these vectors to 0 and generate the sample of the counterfactual variable $Y_{\mathbf{X_S}=\mathbf{0}_{|\mathbf{S}|}}$ to replace y in them. Then, let $\mathcal{B} = \{v = (\mathbf{x_S}, \mathbf{x}_{\mathbf{V}\setminus\mathbf{X_S}}, y_{\mathbf{X_S}=\mathbf{0}_{|\mathbf{S}|}}) \in \mathcal{A} \mid y_{\mathbf{X_S}=\mathbf{0}_{|\mathbf{S}|}} = 0 \text{ occurs in } v\}$, where $\mathbf{V}$ is the set of all variables except Y. Hence, the Change Rate can be calculated by $|\mathcal{B}|/N$. In fact, the set $\mathcal{B}$ pick up a group of people that given $\mathbf{X_S} = \mathbf{x_S}$ and $Y = 1$, when we change the value of $\mathbf{X_S}$ to zero, then the value of Y changes to 0 from 1. Therefore, the Change Rate calculates the proportion of people in set $\mathcal{B}$ to the total population, which is an estimate of the counterfactual probability $\mathrm{P}(Y_{\mathbf{X_S}=\mathbf{0}_{|\mathbf{S}|}} = 0 \mid \mathbf{V} = \mathbf{v}, Y = 1)$. The higher the proportion of this kind of people, the greater the impact of $\mathbf{X_S}$ on Y. We repeat the above experiment ten times with different seeds and take the average as the result, as shown in Table 1.

CR1 shows the average change rate of different methods for attributing one cause, while numbers in parentheses in the second line are their standard deviations. Recall that when only one cause is attributed, CCCE degenerates into POSTTCE. It can be seen that POSTTCE and CCCE perform much better than other methods and have similar change rates, 0.6636 and 0.6713, respectively. In addition, CR2 in Table 1 shows the average change rate for attributing two causes. Due to monotonicity, all methods perform better

Table 2: Change rate of methods for attribution with interaction effect.

	Rand	Post	PN	PS	PNS	ACE	PostTCE	CCCE
CR2	0.3976	0.5863	0.5300	0.5263	0.5340	0.5471	0.6088	**0.6597**
std.	(0.0206)	(0.0266)	(0.0263)	(0.0257)	(0.0320)	(0.0310)	(0.0238)	(0.0196)

Table 3: Change rate of methods for attribution under different level of monotonicity.

p		Rand	Post	PN	PS	PNS	ACE	PostTCE	CCCE
0	CR1	0.1795	0.5375	0.5342	0.5337	0.5352	0.5348	0.6517	**0.6623**
	std.	(0.0098)	(0.0126)	(0.0163)	(0.0140)	(0.0165)	(0.0124)	(0.0171)	(0.0172)
	CR2	0.3080	0.6475	0.5363	0.5375	0.5446	0.6438	0.6643	**0.6748**
	std.	(0.0110)	(0.0743)	(0.0096)	(0.0127)	(0.0105)	(0.0690)	(0.0170)	(0.0119)
0.2	CR1	0.1772	0.5154	0.5119	0.5210	0.5116	0.5100	0.6345	**0.6413**
	std.	(0.0096)	(0.0107)	(0.0060)	(0.0164)	(0.0076)	(0.0121)	(0.0170)	(0.0187)
	CR2	0.3032	0.6293	0.5098	0.5110	0.5126	0.6214	0.6437	**0.6541**
	std.	(0.0089)	(0.0722)	(0.0081)	(0.0155)	(0.0169)	(0.0748)	(0.0192)	(0.0169)
0.4	CR1	0.1755	0.4942	0.4902	0.4920	0.4915	0.4871	0.6177	**0.6203**
	std.	(0.0096)	(0.0183)	(0.0150)	(0.0098)	(0.0112)	(0.0146)	(0.0200)	(0.0126)
	CR2	0.2852	0.6047	0.4893	0.4908	0.4888	0.6068	0.6173	**0.6348**
	std.	(0.0089)	(0.0722)	(0.0081)	(0.0155)	(0.0169)	(0.0748)	(0.0192)	(0.0169)
0.6	CR1	0.1651	0.4698	0.4758	0.4688	0.4675	0.4668	0.5977	**0.5946**
	std.	(0.0090)	(0.0122)	(0.0152)	(0.0113)	(0.0149)	(0.0131)	(0.0208)	(0.0111)
	CR2	0.2786	0.5821	0.4681	0.4716	0.4667	0.5841	0.5941	**0.6101**
	std.	(0.0134)	(0.0701)	(0.0143)	(0.0136)	(0.0161)	(0.0719)	(0.0198)	(0.0103)

than attributing one cause, and PostTCE and CCCE are still much better than others. Besides, CR2 of CCCE is only about 3.72% higher than that of PostTCE, as counterfacual probabilities of all causes are generated independently and randomly, and there is no interaction effect in this case. Therefore, there is little difference between attributing two variables simultaneously and attributing one variable twice.

Next, we use the method of Equation (16) to generate the probabilities of counterfactuals with explicit interaction effect. In particular, $\gamma_1 \sim \text{Uniform}(1, 3)$; $\gamma_Y, \gamma_j \sim \text{Uniform}(-2, 2), j = 2, \ldots, 6$; $\alpha_{ij} \sim \text{Uniform}(0, 0.5)$; $\alpha_{Y1} \sim \text{Uniform}(0.6, 1)$; $\alpha_{Yj} \sim \text{Uniform}(0, 0.5), j = 2, 3$; and $\beta = 2$. We use the same way to calculate the change rate for attributing two causes. The results are shown in Table 2.

In this case, there is an interaction effect between X_2 and X_3. It can be seen that the change rate of CCCE (0.6597) is the highest, at least 5% higher than other methods. This is because when there is the interaction effect, only CCCE can attribute multiple causes simultaneously while other methods can only calculate one cause at a time.

As mentioned before, the monotonicity assumption may not be true in practice. Hence, we next test the performance of CCCE and the baselines when that monotonicity assumption does not hold. In this experiment, we randomly select several causal graphs for simulation experiments. The experimental setup is exactly the same as that in Table 1 in our paper, except that when we generate counterfactual samples, a proportion $p = (0, 0.2, 0.4, 0.6)$ of random samples will not meet the monotonicity assumption. We present the result of one of the causal graphs here, as shown in Table 3, while the other causal graphs have similar results.

It can be seen that with the decline of the level of monotonicity, the effects of all methods have declined to a certain extent. However, even so, under different levels of monotonicity, the performance of each method is similar to that in Table 1, that is, the performance of PostTCE and CCCE is much better than that of other methods, while the performance of CCCE is slightly better than that of PostTCE, but the difference is not significant. This is because this experiment is the same as the setup of Table 1, which is conducted without interaction effect. This also shows that even if monotonicity is not satisfied, our method can still achieve very good results. In addition, when the monotonicity assumption is destroyed, the CR2 values of some methods are less than their CR1 values. In other words, the effect of attributing two causes is not as good as that of attributing one cause, which also shows that the monotonicity assumption is indeed destroyed.

Table 4: Change rate of methods for attributing the expression of Mek in Figure 2.

	RAND	POST	PN	PS	PNS	ACE	POSTTCE	CCCE
CR1	0.2276	0.5406	0.5335	0.5344	0.5354	0.5358	0.6219	**0.6252**
std.	(0.0165)	(0.0160)	(0.0132)	(0.0123)	(0.0205)	(0.0161)	(0.0207)	(0.0138)
CR2	0.3356	0.5330	0.5330	0.5325	0.5328	0.5380	0.6215	**0.6790**
std.	(0.0182)	(0.0145)	(0.0211)	(0.0147)	(0.0083)	(0.0203)	(0.0244)	(0.0196)

5.2 REAL DATA

In this experiment, we apply our method to a real world dataset about the expression levels of proteins and phospholipids [Sachs et al., 2005]. The ground truth causal graph has 11 vertices and 19 edges, as shown in Figure 2. We aim to attribute the expression of the variable Mek in Figure 2. Here we only use the observational data with 853 samples for our attribution. We binarize the data by setting the data greater than the median value to 1, and 0 otherwise.

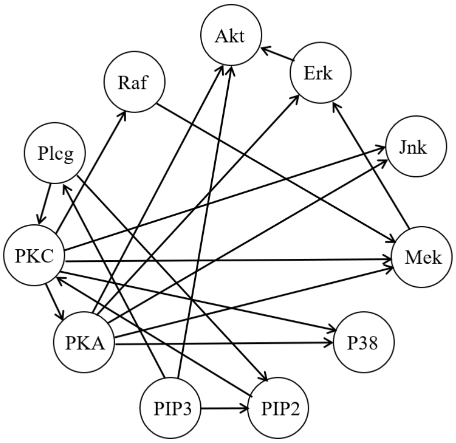

Figure 2: The ground truth causal graph of 11 proteins and phospholipids in [Sachs et al., 2005].

We assume that the causal order is known as $(\mathrm{PKC}, \mathrm{PKA}, \mathrm{Raf}, \mathrm{Jnk}, \mathrm{P38}, \mathrm{Mek}, \mathrm{Erk}, \mathrm{Akt}, \mathrm{Plcg}, \mathrm{PIP2}, \mathrm{PIP3})$. We take the empirical posterior probabilities $\mathrm{P}(X_j = 1 \mid \mathbf{A}_j = \mathbf{a}_j)$ as the counterfactual probabilities $\mathrm{P}((X_j)_{\mathbf{a}_j})$ and calculate the change rates of different methods in the same way as Section 5.1. The results are shown in Table 4.

Similarly, CR1 in Table 4 is the average change rate for attributing one cause. Again, POSTTCE and CCCE outperform other methods. Note that the change rates of POSTTCE and CCCE are close, while the latter has a smaller standard deviation. As for CR2, the average change rate for attributing two causes, CCCE performs best and its change rate is 5.75% higher and a smaller standard deviation compared with POSTTCE. This implies that there is an interaction effect between the causes of Mek. It is worth noting that values of CR2 obtained by POST, PN, PS, PNS and POSTTCE are lower than those of CR1, respectively. We conjecture that this is because the monotonicity assumption may not hold in this dataset. Nevertheless, CCCE still has an increased change rate.

6 CONCLUDING REMARKS

In this paper, we propose CCCE to quantify how possibly the result is attributed to causes. In particular, CCCE allows us to attribute several causes simultaneously and characterize the interaction effect between two or more causes. In addition, CCCE can use the observational data of any variable as evidence, including post-treatment variables and variables may be affected by the outcome. We discuss the identifiability of CCCE, and present the assumptions required and identification equations in different cases. According to these results, we extend the result to the general case, and finally give a general result on th identifiability of CCCE. We apply our method to both synthetic and real world datasets and achieve a significant improvement over the existing methods.

In the present paper, all variables involved are binary, which might be a limitation in practice. We plan to extend CCCE to the case of discrete or categorical variables. Relaxing the assumption of a known causal order is also a future work direction.

Acknowledgements

This work is supported in part by the National Key R&D Program of China (2020YFE0204200) and in part by the National Natural Science Foundation of China under Grant 11971040. Authors would like to thank reviewers for their helpful and constructive suggestions.

References

R. U. Ayres and J. Walter. The greenhouse effect: Damages, costs and abatement. *Environmental and Resource Economics*, 1(3):237–270, 1991.

G. E. Belch and A. Michael. *Advertising and promotion:An integrated marketing communications perspective*. Irwin-McGraw Hill, 1998.

A. P. Dawid. Causal inference without counterfactuals. *Publications of the American Statistical Association*, 95 (450):407–424, 2000.

A. P. Dawid, D. L. Faigman, and S. E. Fienberg. Fitting science into legal contexts: Assessing effects of causes or causes of effects? *Sociological Methods & Research*, 43: 359–390, 2014.

P. Dawid, M. Musio, and S. E. Fienberg. From statistical evidence to evidence of causality. *Bayesian Analysis*, 11: 725–752, 2016.

M. Jensen and R. Ruback. The market for corporate control: The scientific evidence. *Journal of Financial Economics*, 11:5–50, 1983.

T. Jin. Identifying conditional causal effects, 2012.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

C. Lin, S. Venkataraman, and S. D. Jap. Media multiplexing behavior: Implications for targeting and media planning. *Marketing Science*, 32(2):310–324, 2013.

Z. T. Lu, Z. Geng, W. Li, S. Y. Zhu, and J. Z. Jia. Evaluating causes of effects by posterior effects of causes. *Biometrika*, 2022.

P. A. Naik and K. Raman. Understanding the impact of synergy in multimedia communications. *Journal of marketing research*, 40(4):375–388, 2003.

R. Rosenbaum Paul and B. Rubin Donald. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82:669–688, 1995.

J. Pearl. *Causality: Models, reasoning, and inference, second edition*. Cambridge University Press, 2009.

K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

H. A. M. Voorveld and S. M. F. Valkenburg. The fit factor: The role of fit between ads in understanding cross-media synergy. *Journal of Advertising*, 44(3):1–11, 2015.