

## Causal, Casual and Curious

Judea Pearl\*

# Graphoids over Counterfactuals

**Abstract:** Augmenting the graphoid axioms with three additional rules enables us to handle independencies among observed as well as counterfactual variables. The augmented set of axioms facilitates the derivation of testable implications and ignorability conditions whenever modeling assumptions are articulated in the language of counterfactuals.

**Keywords:** conditional independence, counterfactuals, graphoids, testable implications, consistency axiom, ignorability, mediation

DOI 10.1515/jci-2014-0028

## 1 Motivation

Consider the causal Markov chain  $X \rightarrow Y \rightarrow Z$  which represents the structural equations:

$$y = f(x, u_1) \quad (1)$$

$$z = g(y, u_2) \quad (2)$$

with  $u_1$  and  $u_2$  being omitted factors such that  $X, u_1, u_2$  are mutually independent.

It is well known that, regardless of the functions  $f$  and  $g$ , this model implies the conditional independence of  $X$  and  $Z$  given  $Y$ , written as

$$X \perp\!\!\!\perp Z \mid Y \quad (3)$$

This can be readily derived from the independence of  $X, u_1$ , and  $u_2$ , and it also follows from the  $d$ -separation criterion, since  $Y$  blocks all paths between  $X$  and  $Z$ .

However, the causal chain can also be encoded in the language of counterfactuals by writing

$$Y_x(u) = f(x, u_1) \quad (4)$$

$$Z_{xy}(u) = g(y, u_2) = Z_y(u) \quad (5)$$

where  $u$  stands for all omitted factors (in our case  $u = \{u_1, u_2\}$ ) and  $Y_x(u)$  stands for the value that  $Y$  would take in unit  $u$  had  $X$  been  $x$ . Accordingly, the functional and independence assumptions embedded in the chain model translate into the following counterfactual statements:

$$Z_{xy} = Z_y \quad (6)$$

$$X \perp\!\!\!\perp Y_x \quad (7)$$

$$Z_{xy} \perp\!\!\!\perp (Y_x, X) \quad (8)$$

---

\*Corresponding author: Judea Pearl, Department of Computer Science, University of California – Los Angeles, Los Angeles, CA 90095-1596, USA, E-mail: judea@cs.ucla.edu

Equation (6) represents the missing arrow from  $X$  to  $Z$ , while eqs (7) and (8) convey the mutual independence of  $X$ ,  $u_1$ , and  $u_2$ .<sup>1</sup>

Assume now that we are given the three counterfactual statements (6)–(8) as a specification of some uncharted model; the question arises: Are these statements testable? In other words, is there a statistical test conducted on the observed variables  $X$ ,  $Y$ , and  $Z$  that could prove the model wrong? On the one hand, none of the three defining conditions (6)–(8) is testable in isolation, because each invokes a counterfactual entity. On the other hand, the fact that the chain model of eqs (1) and (2) yields the conditional independence of eq. (3) implies that the combination of all three counterfactual statements should yield a testable implication.

This paper concerns the derivation of testable conditions like eq. (3) from counterfactual sentences like eqs (6)–(8). Whereas graphical models have the benefits of inferential tools such as  $d$ -separation [1, 2, p. 335] for deriving their testable implications, counterfactual specifications must resort to the graphoid axioms,<sup>2</sup> which, on their own, cannot reduce subscripted expressions like eqs (6)–(8) into a subscript-free expression like eq. (3). To unveil the testable implications of counterfactual specifications, the graphoid axioms must be supplemented with additional inferential machinery.

We will first prove that eq. (3) indeed follows from eqs (6) to (8) and then tackle the general question of deriving testable sentences from any given collection of counterfactual statements of the conditional independence variety. To that end, we will augment the graphoid axioms with three auxiliary inference rules, which will enable us to remove subscripts from variables and, if feasible, derive sentences in which all variables are unsubscripted, that is, testable. These auxiliary rules will rely on the composition axiom [1, p. 229]

$$X_w = x \implies Y_{xw} = Y_w \quad (9)$$

which was shown to be sound and complete relative to recursive models [6, 7].<sup>3</sup> In the special case of  $W = \{\emptyset\}$  the axiom is known as *consistency rule*:

$$X = x \implies Y_x = Y \quad (10)$$

and is discussed by Robins [9] and Pearl [10].

## 2 Deriving testables from non-testables

In this section we will show that eq. (3) can be derived from eqs (6) to (8) with the help of eq. (9).

We first note that substituting eq. (6) into eq. (8) yields

$$Z_y \perp\!\!\!\perp (Y_x, X) \quad (11)$$

which is a universally quantified formula, stating that for all  $z$ ,  $y$ ,  $y'$ ,  $x$ ,  $x'$  in the respective domains of  $Z$ ,  $Y$ , and  $X$ , the following independence condition holds

$$Z_y = z \perp\!\!\!\perp (Y_x = y', X = x') \quad (12)$$

We next note that, for the special case of  $x' = x$ , eq. (12) yields

$$Z_y = z \perp\!\!\!\perp (Y_x = y', X = x)$$

<sup>1</sup> Rules for translating graphical models to counterfactual notation are given in Pearl [1, pp. 232–4], based on the structural semantics of counterfactuals. The rules represent the omitted factors affecting any variable, say  $Y$ , by the set of counterfactuals  $Y_{pa(Y)}$ , where  $pa(Y)$  stands for the parents of  $Y$  in the diagram.

<sup>2</sup> The graphoid axioms are axioms of conditional independence, first formulated by Dawid [3] and Spohn [4]. Their connections to graph connectivity and to other notions of “information relevance” were established by Pearl and Paz [5] and are described in detail in Pearl [1, pp. 78–133, 2, p. 11].

<sup>3</sup> The axiom of “composition” was first stated in Holland [8, p. 968]. Its completeness rests on a few technical conditions such as uniqueness and effectiveness [7].

or, using eq. (10)

$$Z_y = z \perp\!\!\!\perp (Y = y', X = x) \text{ for all } y, z, y', x \quad (13)$$

This can be written succinctly as

$$Z_y \perp\!\!\!\perp (Y, X) \quad (14)$$

Our next task is to remove the subscript from  $Z_y$ . This is done in two steps. First we apply the graphoid rule of “weak union” (i.e.,  $W \perp\!\!\!\perp (V, S) \implies W \perp\!\!\!\perp V \mid S$ , [1, p. 11]) to obtain

$$Z_y \perp\!\!\!\perp (Y, X) \implies Z_y \perp\!\!\!\perp X \mid Y \quad (15)$$

Second, we explicate the components of eq. (15) and write

$$Z_y \perp\!\!\!\perp (X, Y) \implies Z_y = z \perp\!\!\!\perp X = x \mid Y = y' \quad (16)$$

for all  $y, z, x$ , and  $y'$ . Again, for the special case of  $y' = y$ , eq. (10) permits us to remove the subscript from  $Z_y$  and write

$$Z = z \perp\!\!\!\perp X = x \mid Y = y \quad \text{for all } x, y, z \quad (17)$$

Finally, since the last independency holds for all  $x, y$ , and  $z$ , we can write it in succinct notation as

$$Z \perp\!\!\!\perp X \mid Y$$

which is subscript free and coincides with the testable implication of eq. (3).

To summarize, we have shown that the subscripts in eq. (11) can be removed in two steps. First

$$Z_y \perp\!\!\!\perp (Y_x, X) \implies Z_y \perp\!\!\!\perp (Y, X) \quad (18)$$

and second,

$$Z_y \perp\!\!\!\perp (Y, X) \implies Z \perp\!\!\!\perp X \mid Y \quad (19)$$

Moreover, we see that eq. (3) follows from eq. (8) alone and does not require the exogeneity assumption expressed in eq. (7).

### 3 Augmented graphoid axioms

In this section, we will identify three general rules that, when added to the graphoid axioms, will enable us to derive testable implications without referring back to the consistency axiom of eq. (10). The three rules are as follows:

#### Rule 1

$$V \perp\!\!\!\perp (X_w, Y_{xw}, S) \mid R \implies V \perp\!\!\!\perp (X_w, Y_w, S) \mid R \quad (20)$$

#### Rule 2

$$V \perp\!\!\!\perp R \mid (X_w, Y_{xw}, S) \implies V \perp\!\!\!\perp R \mid (X_w, Y_w, S) \quad (21)$$

#### Rule 3

$$V \perp\!\!\!\perp (Y_{xw}, S) \mid (X_w, R) \implies V \perp\!\!\!\perp (Y_w, S) \mid (X_w, R) \quad (22)$$

Rules 1 and 2 state that a subscript  $x$  can be removed from  $Y_{xw}$  whenever  $Y_{xw}$  stands in conjunction with  $X_w$ , be it before or after the conditioning bar. In our example we had  $W = \{\emptyset\}$ . Rule 3 states that a subscript  $x$  can be removed from  $Y_{xw}$  whenever  $X_w$  appears in the conditioning set. The symbols  $V, S, R$  in eqs (20)–(22) stand for any set of variables, observable as well as counterfactual.

The proof of these three rules follows the path that led to the derivation of eqs (18) and (19).

For mnemonic purposes we can summarize these rules using the following shorthand:

**Rule 1–2**

$$(X_w, Y_{xw}) \implies (X_w, Y_w) \quad (23)$$

**Rule 3**

$$(Y_{xw} | X_w) \implies (Y_w | X_w) \quad (24)$$

## 4 Deriving ignorability relations

Unveiling testable implications is only one application of the augmented graphoid axioms in Section 3. Not less important is the ability of these axioms to justify ignorability relations which a researcher may need for deriving causal effect estimands [1, 8, 11, 12].<sup>4</sup>

Consider the sentence  $Z_x \perp\!\!\!\perp (Y_z, X)$  which may be implied by a certain process and assume we wish to estimate the causal effect of  $Z$  on  $Y$ ,  $P(Y_z = y)$  from non-experimental data. For this estimation to be unbiased, the conditional ignorability  $Z \perp\!\!\!\perp Y_z | W$  needs to be assumed, where  $W$  is some set of observed covariates. Using Axiom (22) we can show that  $W = X$  satisfies the ignorability assumptions and, therefore, adjustment for  $X$  will yield a bias-free estimate of the causal effect  $P(Y_z = y)$ . This can be shown as follows:

$$Z_x \perp\!\!\!\perp (Y_z, X) \implies Z_x \perp\!\!\!\perp Y_z | X$$

(using the graphoid rule of “weak union”) and by Rule (22) we obtain

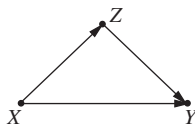
$$Z_x \perp\!\!\!\perp Y_z | X \implies Z \perp\!\!\!\perp Y_z | X$$

We therefore can write

$$\begin{aligned} P(Y_z = y) &= \sum_x P(Y_z = y | X = x)P(X = x) \\ &= \sum_x P(Y_z = y | Z = z, X = x)P(X = x) \\ &= \sum_x P(Y = y | Z = z, X = x)P(X = x). \end{aligned} \quad (25)$$

Equation (25) is none other but the standard adjustment formula for the causal effect of  $Z$  on  $Y$ , controlling for  $X$ .

The process can also be reversed; we start with a needed, yet unsubstantiated ignorability condition, and we ask whether it can be derived from more fundamental conditions which are either explicit in the model or defensible on scientific grounds. Consider, for example, an unconfounded mediation model in which treatment  $X$  is randomized and assume we seek to estimate to effect of the mediator  $Z$  on the outcome  $Y$ . (The model is depicted in Figure 1.) Operationally, we know that the ignorability condition  $Z \perp\!\!\!\perp Y_z | X$  would allow



**Figure 1** Unconfounded mediation model implying the conditional ignorability  $Z \perp\!\!\!\perp Y_z | X$

<sup>4</sup> Reliance on the assumptions of conditional ignorability [8, 11, 12], which are cognitively formidable, is one of the major weaknesses of the potential outcome framework [1, pp. 350–1]. Axioms (20)–(22) permit us to derive needed ignorability conditions from other counterfactual statements which are perhaps more transparent.

us to obtain the desired effect  $P(Y_z = y)$  by adjusting for  $X$ , as shown in the derivation of eq. (25). However, lacking graphs for guidance, it is not clear whether this condition follows from the assumptions embedded in the model; a formal proof is therefore needed. The assumptions explicit in the model take the form

- (a)  $X_z = X$
- (b)  $X \perp\!\!\!\perp (Z_x, Y_{zx})$
- (c)  $Z_x \perp\!\!\!\perp Y_{zx}$ .

(a) states that  $Z$  does not affect  $X$ , (b) represents the assumption that  $X$  is randomized, and (c) stands for the no-confounding assumption, that is, all factors affecting  $Z$  when  $X$  is held constant are independent of those affecting  $Y$  when  $X$  and  $Z$  are held constant [1, p. 232, 343]. These factors stand precisely for the “error terms” that enter the structural equations for  $Z$  and  $Y$ , respectively; hence, they have clear process-based interpretations and avail themselves to plausibility judgments.

To show that the desired ignorability condition  $Z \perp\!\!\!\perp Y_z | X$  follows from (a), (b) and (c), we can use Rule 3 (eq. 22) as follows. First, the standard graphoid axioms dictate

$$X \perp\!\!\!\perp (Z_x, Y_{zx}) \text{ and } Z_x \perp\!\!\!\perp Y_{zx} \implies Z_x \perp\!\!\!\perp Y_{zx} | X$$

Next, applying Rule 3 twice, together with  $X = X_z$ , gives

$$Z_x \perp\!\!\!\perp Y_{zx} | X \implies Z \perp\!\!\!\perp Y_{zx} | X \implies Z \perp\!\!\!\perp Y_{zx} | X_z \implies Z \perp\!\!\!\perp Y_z | X_z \implies Z \perp\!\!\!\perp Y_z | X$$

which yields the desired ignorability condition.

These derivations can be skipped, of course, when we have a graphical model for guidance. The adjustment formula (25) could then be written by inspection, since  $X$  satisfies the back-door condition relative to  $Z \rightarrow Y$ . However, researchers who mistrust graphs and insist on doing the entire analysis by algebraic methods would need to use Rules 1–3 to justify the ignorability condition from assumptions (a), (b) and (c).

## 5 Conclusions

Rules 1–3, when added to the graphoid axioms, allow us to process conditional-independence sentences involving counterfactuals and derive both their testable implications and implications that are deemed necessary for identifying causal effects. We conjecture that Rules 1–3 are *complete* in the sense that all implications derivable from the graphoid axioms together with the consistency rule (18) are also derivable using the graphoid axioms together with Rules 1–3.

Augmented graphoids are by no means a substitute for causal diagrams, since the complexity of finding a derivation using graphoid axioms may be exponentially hard [13]. Diagrams, on the other hand, offer simple graphical criteria (e.g.,  $d$ -separation or back-door) for deriving testable implications and effect estimands. In reasonably sized problems, these criteria can be verified by inspection, while, in large problems, they can be computed in polynomial time [14, 15]. The secret of diagrams is that they embed all the graphoid axioms in their structure and, in effect, pre-compute all their ramifications and display them in graphical patterns.

**Acknowledgment:** Sander Greenland and Jin Tian provided helpful comments on an early version of this note. This research was supported in parts by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211 and #IIS-1018922, and ONR #N000-14-09-1-0665 and #N00014-10-1-0933.

## References

1. Pearl J. Causality: models, reasoning, and inference, 2nd ed. New York: Cambridge University Press, 2009.
2. Pearl J. Probabilistic reasoning in intelligent systems. San Mateo, CA: Morgan Kaufmann, 1988.
3. Dawid A. Conditional independence in statistical theory. J R Stat Soc Ser B 1979;41:1–31.

4. Spohn W. Stochastic independence, causal independence, and shieldability. *J Philos Logic* 1980;9:73–99.
5. Pearl J, Paz A. GRAPHOIDS: a graph-based logic for reasoning about relevance relations. In: Boulay BD, Hogg D, Steels L, editors. *Advances in artificial intelligence-II*. North-Holland, 1987:357–63.
6. Galles D, Pearl J. [An axiomatic characterization of causal counterfactuals](#). *Found Sci* 1998;3:151–82.
7. Halpern J. Axiomatizing causal reasoning. In: Cooper G, Moral S, editors. *Uncertainty in artificial intelligence*. San Francisco, CA: Morgan Kaufmann, 1998:202–10. Also, *J Artif Intell Res* 2000;12:17–37.
8. Holland P. [Statistics and causal inference](#). *J Am Stat Assoc* 1986;81:945–60.
9. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Math Model* 1986;7:1393–512.
10. Pearl J. On the consistency rule in causal inference: an axiom, definition, assumption, or a theorem? *Epidemiology* 2010;21:872–5.
11. Rosenbaum P, Rubin D. [The central role of propensity score in observational studies for causal effects](#). *Biometrika* 1983;70:41–55.
12. Rubin D. [Estimating causal effects of treatments in randomized and nonrandomized studies](#). *J Educ Psychol* 1974;66:688–701.
13. Geiger D. Graphoids: a qualitative framework for probabilistic inference. PhD thesis, Department of Computer Science, University of California – Los Angeles, Los Angeles, CA, 1990.
14. Shpitser I, Pearl J. Complete identification methods for the causal hierarchy. *J Mach Learn Res* 2008;9:1941–79.
15. Tian J, Paz A, Pearl J. Finding minimal separating sets. Technical report R-254, University of California – Los Angeles, Los Angeles, CA, 1998.