Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

# The Blessings of Multiple Causes: Rejoinder

Yixin Wang & David M. Blei

**To cite this article:** Yixin Wang & David M. Blei (2019) The Blessings of Multiple Causes: Rejoinder, Journal of the American Statistical Association, 114:528, 1616-1619, DOI: 10.1080/01621459.2019.1690841

**To link to this article:** https://doi.org/10.1080/01621459.2019.1690841

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 528, 1616–1619: Rejoinder
https://doi.org/10.1080/01621459.2019.1690841



# The Blessings of Multiple Causes: Rejoinder

Yixin Wang[a] and David M. Blei[a,b]

[a]Department of Statistics, Columbia University, New York, NY; [b]Department of Computer Science, Columbia University, New York, NY

We thank all the discussants for taking the time and energy to build on this work; and we thank the editors for putting together an engaging and thought-provoking collection of discussions. After reading these contributions, we were struck that these are not mere discussions—indeed, each is an article in itself. This collection pushes forward "The Blessings of Multiple Causes" in significant ways, offering new theory, new criticism, and new application. After highlighting some of the themes of these articles, we will turn to each individually.

"The Blessings of Multiple Causes" provide assumptions, theory, and algorithms for multiple causal inference. The deconfounder algorithm involves modeling the causes, using the model to infer a substitute confounder, and then using the substitute confounder in a downstream causal inference. The deconfounder is not a black-box solution to causal inference. Rather, it is a way to use careful domain-specific modeling in the service of causal inference.

Causal inference with the deconfounder involves a number of assumptions and trade-offs, and many of the discussants highlighted these. Among them are the following. (1) There can be no unobserved single-cause confounders. (2) When we apply the deconfounder, we trade an increase in estimation variance for a reduction in confounding bias; there is no free lunch. (3) We do not recommend using the deconfounder with causally dependent causes, such as a time series; finding a substitute confounder may be too difficult in these scenarios.

There are many directions for further research, and the discussants have pointed out several of the most important ones. We need a more complete picture of identification; D'Amour (2019) and the discussions here make good progress (see Table1). We need to understand the finite-sample properties of the deconfounder, and how to estimate uncertainty about causal inferences when using a substitute multi-cause confounder. We need rigorous methods of model criticism for assessing the validity of the substitute confounder.

Deconfounder-like methods have already been used for genome-wide association studies (e.g., Pritchard et al. 2000) and estimating peer effects in networks (Shalizi and McFowland III 2016). More broadly, the deconfounder strategy points to many applications, including in genetics, psychology, education, and marketing, where factor models are routinely fit to large-scale data. We hope that statisticians and machine learners will continue to study multiple causality, and that scientists and other practitioners will adapt the deconfounder to help analyze and understand their observational data.

## 1. Athey, Imbens, and Pollmann

Athey, Imbens, and Pollmann (AIP) consider a problem in economics: how do the prices of products affect their demand? The causes are prices; the outcome is demand; and the unobserved confounders are shocks to demand that also affect price. AIP apply the deconfounder to a setting of two products, gracefully using their domain expertise to directly construct a substitute confounder. They show that the deconfounder only helps when the two products have highly correlated demand shocks, that is, when there is shared unobserved confounding. AIP's application beautifully illustrates the importance of domain knowledge to the deconfounder.

AIP compare two methods for estimating causal effects. The deconfounder uses the average price of the two products as the substitute confounder; the instrumental variable approach uses the price of one product as an instrument for the other. It is interesting that these two strategies work best in opposite cases. The deconfounder works when the demand shocks of the two products are highly correlated. The instrumental variable works when they are not (or weakly) correlated. More precisely, both the deconfounder and the instrumental variable approach require the prices be correlated. But the deconfounder requires that the driver of this correlation also affects the outcome, while the instrumental variable approach requires that it not affect the outcome.

AIP's method further suggests extending the deconfounder to more general structures of shared confounding. Unlike the simpler settings in the article, AIP examine multiple causal problems: each product's price affects a different outcome, but with shared unobserved confounders. We imagine that other scientific settings bear the same parallel structure.

## 2. Imai and Jiang

Imai and Jiang (IJ) discuss two technical aspects of the deconfounder.

**CONTACT** Yixin Wang yixin.wang@columbia.edu Department of Statistics, Columbia University, New York, NY.

**Table 1.** Identification in multiple causal inference.

| Causal quantity | Result | Condition | Source |
|---|---|---|---|
| $P(Y(\boldsymbol{a}))$ | Non-ID | No conditions | D'Amour (2019) |
| $\mathbb{E}[Y(\boldsymbol{a})] - \mathbb{E}[Y(\boldsymbol{a}')]$ | ID | Consistent substitute confounder; Categorical substitute confounder; No confounder/cause interaction; Differentiable relationships | Theorem 6 (WB) |
| $\mathbb{E}_A[\mathbb{E}_Y[Y(a_{1:k}, A_{(k+1):m})]]$ | ID | Consistent substitute confounder; $A_{1:k}$ satisfy overlap | Theorem 7 (WB) |
| $\mathbb{E}[Y(\boldsymbol{a}') \mid \boldsymbol{A} = \boldsymbol{a}]$ | ID | Consistent substitute confounder; $\boldsymbol{a}'$ and $\boldsymbol{a}$ map to same substitute | Theorem 8 (WB) |
| $\mathbb{E}[Y(\boldsymbol{a})]$ | ID | $\mathbb{E}[U \mid \mathbf{A}]$ nonlinear; $\mathbb{E}[Y \mid \mathbf{A}, U]$ linear | Section 2.1 (IJ) |
| $\mathbb{E}[Y(\boldsymbol{a})]$ | ID | Measure instrument $W$; Instrument $W$ satisfies overlap | Section 2.2 (IJ) |
| $\int Y(\boldsymbol{a}) q_1(\boldsymbol{a})\, \mathrm{d}\boldsymbol{a} \quad - \int Y(\boldsymbol{a}) q_2(\boldsymbol{a})\, \mathrm{d}\boldsymbol{a}$ | ID | $p(\boldsymbol{a} \mid z) > 0$ when $q_1(\boldsymbol{a}), q_2(\boldsymbol{a}) > 0$ | Section 2.3 (IJ) |

NOTE: ID = identifiable.

They first point out the difficulty of defining "multi-cause" and in particular of defining the assumption "no unobserved single-cause confounders." In the DAG language, this assumption requires (1) the causal graph resides in a class where unobserved confounders must be parents of two or more causes and (2) the causal problem be faithful to the graph (Spirtes 2010).

We agree with IJ that it is difficult to simultaneously express such graphical and faithfulness conditions in the potential outcomes notation. The definition in the article attempts to express faithfulness by considering the smallest sigma algebra that renders the causes conditionally independent (see condition 2 in Definition 4). Note this definition excludes those multi-cause confounders that can be separated into multiple single-cause confounders, as illustrated in Figure 2 of IJ's article.

IJ correctly note that it is unclear whether the definition of "no unobserved single-cause confounders" in the article is equivalent to the one we intended in the DAG language. As suggested by IJ at JSM 2019, a more precise form of "no unobserved single-cause confounders" may be: *there exist a random variable $Z$ s.t. (1) $Z$ satisfies $p(\boldsymbol{a} \mid z) = \prod_{j=1}^{m} p(a_j \mid z)$ and no sigma-algebra smaller than $\sigma(Z)$ satisfies this equation; (2) $A_1, \ldots, A_m \perp Y(\boldsymbol{a}) \mid Z$.* Moreover, assessing the credibility of "no unobserved single-cause confounders" may require substantial domain expertise. How to rigorously translate graphical and faithfulness conditions into the potential outcomes notation is an interesting direction of research.

The second thread of IJ's article is about causal identification of a complete intervention $\mathbb{E}[Y(\boldsymbol{a})]$, and the difficulty of the deconfounder in satisfying overlap. Because the substitute confounder $Z$ is a function of the causes $\boldsymbol{A}$, the overlap condition $P(\boldsymbol{A} \mid Z) > 0$ can be stringent. IJ consider three ways forward: parametric assumptions, instrumental variables, and stochastic interventions.

IJ's parametric approach achieves the identification of $\mathbb{E}[Y(\boldsymbol{a})]$ by the incongruence between the linear outcome model and the nonlinear factor model. Related to IJ's setting, Theorem 6 in the article achieves identification via the incongruence between the differentiability of the outcome model and the non-differentiability of the substitute confounder. IJ's result and Theorem 6 suggest that the idea of incongruence may serve as a general approach to causal identification.

IJ's instrumental variable approach requires an overlap condition that is weaker than the one required by the deconfounder. But, as IJ illustrate in their discrete-variable example, this overlap condition may become more stringent as the number of causes increases. Notice there may be an increasing number of instrumental variables as the number of causes increases, though not one of them might satisfy overlap by itself. IJ's thinking is suggestive of a direction of future investigation: how to combine multiple instrumental variables in multi-cause problems to satisfy overlap and obtain causal identification.

The final approach IJ explore is stochastic intervention. It tackles the problem of overlap by restricting the causal queries. This approach relates to Theorem 8 in the article, which restricts the causal queries to those interventions that map to the same value of the substitute confounder. But IJ's approach is more powerful than Theorem 8 because it handles causally dependent causes. We look forward to more developments in the stochastic intervention approach of multiple causal inference.

Including IJ's new results, Table1 summarizes the current landscape of identification results in multiple causality.

## 3. D'Amour

In both his discussion here and his earlier article (D'Amour 2019), Alex D'Amour has significantly contributed to the understanding of multi-causal identification. We have enjoyed a productive conversation with him over the past years. We were glad to read that the feeling is mutual.

In his discussion, D'Amour articulates the fundamental tension between using the causes to infer unobserved confounding and using them to estimate causal effects. In other words, the deconfounder does not provide free lunch: the more information is baked into estimating the substitute confounder, the less information is available for estimating causal quantities. Moreover, the assumption that we can pinpoint the substitute confounder is at odds with "all-cause" overlap, that is, that $P(\boldsymbol{A} \mid Z) > 0$. As D'Amour (2019) points out, both cannot be simultaneously satisfied.

Theorems 6–8 in the article live at one extreme of this tension. They assume a pinpointed substitute confounder and forgo overlap on all the causes. (Note it is still possible for subsets of the causes to satisfy overlap, as in Theorem 7.) The pinpointed substitute confounder is achievable thanks to the multiplicity of the causes and the consistent estimability of factor models. Going forward, how does identification fare as we move away from this extreme? Point identification might no longer be possible, but partial identification might be.

With the same assumptions as Theorems 6–8, D'Amour studies both parametric and nonparametric identification. The parametric direction is risky without strong prior knowledge. But certain applications enjoy parametric models that are worth studying. For example, when we believe causal effects are small, a structural model that is linear in the causes but nonparametric in the unobserved confounder may be reasonably close to reality, $Y = \sum_{j=1}^{m} \beta_j A_j + g(U) + \varepsilon$. Identification conditions for such parametric models can be convenient for practical applications.

In the nonparametric direction, D'Amour explores Theorems 7 and 8 of the article. D'Amour's Proposition 1 summarizes well the essence of the theorems. Toward a more cautious application of the deconfounder, he suggests performing conditional independence tests or sensitivity analysis. This is an important direction of investigation and could be useful in many scientific domains.

## 4. Ogburn, Shpitser, and Tchetgen Tchetgen

Ogburn, Shpitser, and Tchetgen Tchetgen (OSTT) provide a technical meditation on some of the theoretical aspects of the article, and a dissenting opinion about its value. Among their remarks, they claim that there are "foundational errors" with the work and that the "premise is incorrect." These claims are not substantiated. There are no foundational errors; the premise is correct.

The identification results in Theorems 6–8 capitalize on two requirements: (1) the distribution of the causes $p(\boldsymbol{a})$ can be described by a factor model and (2) the factor model pinpoints the substitute confounder $Z$, that is, $Z \overset{\text{a.s.}}{=} f_\theta(\mathbf{A})$ for some $f_\theta$. The first requirement relies on the successful execution of the deconfounder, that is, finding a factor model that captures $p(\boldsymbol{a})$. The conditional independence structure of factor models guarantees that the substitute confounder $Z$ pick up all multi-cause confounders and no multi-cause mediators or colliders. The second requirement is the "consistency of the substitute confounder." It is satisfied when the number of causes goes to infinity and $Z$ remains finite-dimensional. From Lemma 4, it guarantees that $Z$ cannot pick up single-cause confounders, mediators, or colliders.

OSTT's main concern revolves around Lemma 4, which states the substitute confounder cannot pick up information about multi-cause mediators, single-cause mediators, or any of the other graphs that OSTT put forward. Lemma 4 is correct, as is the proof in the article. But Lemma 4 might also seem surprising. Here is an alternative proof.

*Restatement of Lemma 4.* No post-treatment variable can be measurable with respect to a consistent substitute confounder.

*Proof.* First, the substitute cannot pick up any multi-cause post-treatment variables. Otherwise, the substitute cannot render all the causes conditionally independent.

The substitute also cannot pick up any single-cause variables. These variables include pretreatment variables, such as single-cause confounders, and single-cause post-treatment variables, such as single-cause mediators or colliders.

The key idea behind the proof is the following. We assume the causes pinpoint the substitute confounder $Z \overset{\text{a.s.}}{=} f(\boldsymbol{A}; \theta)$, as is the case where there are many causes. The deconfounder further requires that the converse is not true, that is, that the substitute does not pinpoint the causes. This fact holds in a probabilistic model of the causes, such as when the dimension of the substitute stays fixed as the number of causes increases. Further, the deconfounder requires that the factor model cannot have one component of the substitute *a priori* be a deterministic function of another component; this fact also holds in probabilistic factor models. The proof then follows by contradiction: if the substitute picks up single-cause variables then the factor model must be "degenerate," that is, nonprobabilistic.

Here are the details. Suppose the substitute $Z$ does pick up a single-cause variable. Then separate $Z$ into a single-cause component and a multi-cause one, $Z = (Z_s, Z_m)$. Without loss of generality, assume the single-cause component only depends on the first cause. The assumption of a consistent substitute confounder says

$$p(z \mid \boldsymbol{a}, \theta) = p(z_s, z_m \mid \boldsymbol{a}, \theta) = \delta_{(f_s(\boldsymbol{a};\theta), f_m(\boldsymbol{a};\theta))}, \tag{1}$$

where $\boldsymbol{a} = (a_1, \ldots, a_m)$ are the $m$ causes and $f(\cdot)$ are the deterministic functions that map causes to substitute confounders.

Now calculate the conditional distribution of the single-cause component given the causes,

$$\begin{aligned} &p(z_s \mid \boldsymbol{a}) \\ &\quad = p(z_s \mid \boldsymbol{a}, z_m = f_m(\boldsymbol{a}; \theta))), && (2)\\ &\quad = p(z_s \mid a_1, z_m = f_m(\boldsymbol{a}; \theta))), && (3)\\ &\quad = \frac{p(z_s \mid z_m = f_m(\boldsymbol{a}; \theta)) \cdot p(a_1 \mid z_s, z_m = f_m(\boldsymbol{a}; \theta))}{p(a_1 \mid z_m = f_m(\boldsymbol{a}; \theta))}. && (4) \end{aligned}$$

Equation (2) is due to the consistency of substitute confounder. Equation (3) is due to $Z_s \perp A_2, \ldots, A_m \mid A_1, Z_m$. Equation (4) is due to the definition of conditional probability.

Equation (4) and Equation (1) imply that at least one of $p(z_s \mid z_m = f_m(\boldsymbol{a}; \theta))$ and $p(a_1 \mid z_s, z_m = f_m(\boldsymbol{a}; \theta))$ is a point mass. But this is a contradiction: either term being a point mass implies that the factor model is degenerate. The former is a point mass when one component $Z_s$ of the substitute is a deterministic function of another component $Z_m$. The latter is a point mass when the first cause is a deterministic function of the latent $Z$.

Note the same argument would not reach a contradiction for multi-cause variables $Z_m$. The reason is that

$$\begin{aligned} &p(z_m \mid \boldsymbol{a}) \\ &\quad = p(z_m \mid \boldsymbol{a}, z_s = f_s(\boldsymbol{a}; \theta))), && (5)\\ &\quad = \frac{p(a_1, z_m \mid z_s = f_s(\boldsymbol{a}; \theta))) \cdot \prod_{j=2}^{m} p(a_j \mid z_m)}{p(\boldsymbol{a})}, && (6) \end{aligned}$$

where $\prod_{j=2}^{m} p(a_j \mid z_m)$ can converge to a point mass with non-degenerate factor models and $m \to \infty$. □

OSTT also question the random variable on which we used the Kallenberg construction in Lemmas 1 and 2. Definition 3 is the Kallenberg construction we intended, and it involves

potential outcomes (see Equation (38) in the article). Lemmas 1 and 2 link factor models of the causes to their Kallenberg construction and unconfoundedness, thanks to the consistency of the substitute confounder. Such a substitute cannot separate a multi-cause confounder into single-cause confounders, as the one in OSTT's counterexample does. OSTT claim that the article leaves open that Theorem 7 is "vacuous" because the overlap condition may be impossible to satisfy. D'Amour's discussion of the article shows how Theorem 7 can be useful.

Finally, OSTT remark that requiring a pinpointed substitute implies that the unobserved (multi-cause) confounding is effectively observed. Their intuition is correct—the multiplicity of the causes and the consistent estimability of factor models enable us to effectively observe such multi-cause confounding. It is these two features that form the basis of the deconfounder.

## Acknowledgments

We thank Gemma Moran, Aaron Schein, Keyon Vafa, Victor Veitch, and Alex D'Amour for helpful discussions.

## References

D'Amour, A. (2019), "On Multi-Cause Approaches to Causal Inference With Unobserved Counfounding: Two Cautionary Failure Cases and a Promising Alternative," in *Artificial Intelligence and Statistics*, pp. 3478–3486. [1616,1617]

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000), "Association Mapping in Structured Populations," *The American Journal of Human Genetics*, 67, 170–181. [1616]

Shalizi, C. R., and McFowland III, E. (2016), "Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations," arXiv no. 1607.06565. [1616]

Spirtes, P. (2010), "Introduction to Causal Inference," *Journal of Machine Learning Research*, 11, 1643–1662. [1617]