# Causal Interaction and Effect Modification: Same Model, Different Concepts\*

Luke Keele<sup>†</sup>

Randolph T. Stevenson<sup>‡</sup>

February 2, 2017

Word Count: XXXX

#### Abstract

Social scientists use the concept of interactions to study effect dependency. Such analyses can be conducted using standard regression models. However, an interaction analysis may represent either a causal interaction or effect modification. Under causal interaction, the analyst is interested in whether two treatments have differing effects when both are administered. Under effect modification, the analysts investigates whether the effect of a single treatment varies across levels of a baseline covariate. Importantly, the identification assumptions for these two types of analysis are very different. In this paper, we clarify the difference between these two types of interaction analysis. We demonstrate that this distinction is mostly ignored in the political science literature. We conclude with a review of several applications.

<sup>\*</sup>We would like to thank X for research assistance.

<sup>†</sup>Associate Professor, Georgetown University, Washington, D.C. 20057 Email: lk681@georgetown.edu.

<sup>&</sup>lt;sup>‡</sup>Professor, Department of Political Science, Rice University, Houston, TX 77251, Email: stevenso@rice.edu

### 1 Introduction

Interaction analyses are used to study the possibility of effect dependency and are widely used in the empirical literature in political science. Currently there exists a large literature that provides guidelines for the implementation of interaction analyses with statistical models (Franzese and Kam 2009; Berry, Golder, and Milton 2012; Brambor, Clark, and Golder 2006; Clark, Gilligan, and Golder 2006; Braumoeller 2004; Hainmueller, Mummolo, and Xu 2016a). However, within political science, there are few linkages between the literature on interaction analysis and concepts of causal identification. That is, across many parts of the social sciences, researchers have used methods for causal identification, typically within the potential outcomes framework, to better understand the conditions under which data can be used for causal inference. See Keele (2015) for an overview of the literature on causal identification. However, considerable and important work, primarily in biostatistics, has focused on how matters of casual identification affect for interaction analyses (VanderWeele and Robins 2007, 2008; VanderWeele 2015). One key result in the literature on causal interaction is the distinction between causal interactions and effect modification. In the first, analysts are interested in whether the effect of multiple treatments vary together. Under effect modification, interest is focused on whether a single causal effect varies along levels of a baseline covariate or covariates. While both types of studies rely on identical statistical models, the attendant assumptions and interpretation of statistical results differ greatly across the two types of analysis.

In this article, we formally outline the difference between analyses of causal interaction and effect modification. We show how the causal identification assumptions differ across the two types of studies and how this affects interpretation of the results. Next, we conduct a review of the empirical literature in political science to understand how often researchers note this distinction. We then review two examples from the political science literature to demonstrate how the type of interaction analysis matters to matters of causal identification. We conclude with recommendations for practice.

# 2 Statistical Interactions

Interactions are typically articulated within the context of the linear model as means of operationalizing what we call "effect dependency." To that end, one would define an outcome Y, two covariates D and W and write the following linear model:

$$Y = \beta_0 + \beta_1 D + \beta_2 W + \beta_3 DW + \epsilon \tag{1}$$

Effect dependency can then be expressed by taking the marginal effect with respect to either D or W. For example, the following two marginal effects express two different forms of effect dependency:

$$\frac{\partial Y}{\partial D} = \beta_1 + \beta_3 W$$

$$\frac{\partial Y}{\partial W} = \beta_2 + \beta_3 D$$

In the first, the marginal effect of D depends on the value of W, and in the second the marginal effect of W depends on the value of D. At this point, it is important to note, that the statistical model is completely agnostic as to whether the effect of D depends on W or vice versa. That is, the statistical model itself cannot provide any insight into whether effect dependency should be asymmetric.

The advice in the political science literature also tends to be agnostic about the direction of effect dependency. As Franzese and Kam (2009, p. 16) note:

... all interactions are symmetric. Given this logical symmetry, x and z must necessarily both intervene in the other's relationship to y. In this sense, the language of one variable being the intervening or moderating variable and the other being moderated may best be avoided; if an interaction exists, then all variables involved intervene or moderate the other's relationship to y

The remainder of this article argues that this statement is both correct and incorrect. In

particular, we argue that analysts may be interested in either a causal interaction, effect modification or both, and the correctness of the statement above depends on which type of analysis the researcher is interested in. Specifically, as we show next, the distinction depends on the causal identification assumptions.

### 3 Causal Interactions

First, we take up the concept of causal interaction. We provide but a brief overview of this topic, see VanderWeele (2015) for a complete treatment. Above, we were agnostic as to whether we were interested in causal effects, but now we let D and W denote two binary treatments, where when either is zero treatment is absent, but when either is one two distinct treatments are administered. For example in a GOTV field experiment, D might represent being assigned to the condition where one receives a phone call and W might represent being assigned to the condition where one receives a postcard. All units are assigned to both W and D such that some units receive neither, some receive one or the other, and some units receive both. We assume the analyst is interested in the causal effects of both D and W. We leave questions of causal identifiability for the moment but take them up shortly. We assume that Y is binary, and we denote the conditional probability for Y as:  $p_{dw} = P[y = 1|D = d, W = w]$ . While Y is binary in this example, the generalization to a nonbinary Y is straightforward. Following VanderWeele (2015), we can write the effect of a causal interaction between D and W on Y as

$$(p_{11} - p_{00}) - [(p_{10} - p_{00}) + (p_{01} - p_{00})].$$

The first term,  $(p_{11} - p_{00})$ , is the probability that Y is one when both D and W are one with the probability that Y is one when both D and W are zero. The second term,  $(p_{10} - p_{00})$ , is the effect of the treatment D with the effect of the control removed, and the third term,  $(p_{01} - p_{00})$ , is the effect of W with the effect of the control condition removed. This can be rewritten more compactly as

$$p_{11} - p_{10} - p_{01} + p_{00}.$$

This equation expresses the extent to which the effect of the two treatments together exceeds the effect of each treatment individually. If this term is positive the interaction can be described as "super-additive," since the effect of the two treatments together is larger than the individual treatment effects. If this term is negative, the interaction may be described as "sub-additive." Here, the effect of the two treatments together reduces the overall effect more than the separate treatment effects. More generally,  $p_{11} - p_{10} - p_{01} + p_{00} > 0$  indicates that an intervention on W would be larger in the D = 1 group, while  $p_{11} - p_{10} - p_{01} + p_{00} < 0$  implies the effect of an intervention on W would be larger in the D = 0 group. In the GOTV example, we might expect the interaction to be super-additive since the effect of both a phone call and a postcard should be larger than the effect of just a phone call or just a postcard.

These quantities have a direct correspondence to the familiar quantities from regression models. Assume we were to estimate the following linear model with a multiplicative interaction term:

$$Y = \beta_0 + \beta_1 D + \beta_2 W + \beta_3 DW.$$

The correspondence is as follows  $\beta_0 = p_{00}$ ,  $\beta_1 = (p_{10} - p_{00})$ ,  $\beta_2 = (p_{01} - p_{00})$ , and  $\beta_3 = p_{11} - p_{10} - p_{01} + p_{00}$ . Note that if D and W are both binary the model is saturated, and the above model imposes no functional form assumptions.

When investigators are interested in causal interaction, they can understand whether the effect of one treatment depends on the other treatment. However, this may not be the most fruitful way to understand a causal interaction. Instead, the investigator might alternatively view a causal interaction as a study of the joint effect of two treatments. Assume that for both D and W, when they are set to 0 subjects are exposed to placebo, but when they are each set to 1 an active treatment as is the case in the GOTV example. When this is true, the interaction effect  $p_{11} - p_{10} - p_{01} + p_{00}$  represents the joint effect of the two treatments. VanderWeele (2015, ch. 9) demonstrates that the joint effect of both treatments can be decomposed into three components using the following equation:

$$p_{11} - p_{00} = (p_{10} - p_{00}) + (p_{01} + p_{00}) + (p_{11} - p_{10} - p_{01} + p_{00}).$$

In the equation above,  $(p_{10} - p_{00})$  is the effect of D alone,  $(p_{01} + p_{00})$  is the effect of W alone, and  $(p_{11} - p_{10} - p_{01} + p_{00})$  is the joint effect of D and W. Under this form of causal interaction analysis, interest focuses on the joint effect rather than whether one treatment alters the effect of the other treatment.

A study by Brader, Valentino, and Suhat (2008) provide an example of an empirical study that relies on a research design focused on causal interaction. They study the effect of media cues on immigration attitudes. They hypothesize that media messages on the costs of immigration and cues about the race of immigrants will affect attitudes but may also cause political participation. To test their theory, they use a randomized experiment with two treatments. In one treatment, subjects where exposed to a news story that emphasized the costs of immigration. In the second treatment, they varied the picture of a hypothetical immigrant. Under the control condition, the picture was of a white immigrant, but was a picture of a Hispanic immigrant under the treatment. In this design, they include two forms of control. Under one control condition, subjects viewed a story about immigration that emphasized positive benefits with a white immigrant. Under the other control condition, subjects viewed a story unrelated to immigration. This research design is emblematic of a causal interaction.

We might expect that there are separate effects of both the tone of the news story and the racial cue. We could describe this as effect modification, where the effect of tone varies with levels of racial cue or vice versa. More fruitfully, we might view this as a causal interaction where, there are separate effects of tone and race, but we expect the joint effect to be larger. Basic inspection of their data lend support to the causal interaction interpretation. One outcome in the study was a binary measure of whether subjects decided to send their member of Congress an anti-immigration message. In the true control condition, 32% of subjects selected to send an anti-immigration message. Under the first treatment, where subjects viewed a news story with a negative tone about immigration, the percentage of subjects who send a anti-immigration message declined by 2.5% percentage points, but this effect is not statistically significant. Under the first treatment, where subjects viewed a news story with a the picture of a Hispanic immigrant, the percentage of subjects who send a anti-immigration message declined by 4.7% percentage points,

but this effect is not statistically significant either. In the condition, where the subjects viewed a negative story with a picture of an Hispanic immigrant, the percentage of people who sent an anti-immigration message increases relative to the control condition by nearly 14% points. This is a prototypical causal interaction, where the individual treatments have little effect, but the joint treatment has a strong effect. This is known as a "sufficient cause interaction;" where the outcome occurs only if both treatments are present, but the outcome does not occur if just one of the treatments is active(VanderWeele and Robins 2007, 2008). As such, while causal interactions can be interpreted as whether the effect of one treatment depends on the other treatment, in many cases a sufficient cause interaction may be more appropriate.

#### 3.1 Identification Conditions

Thus far, we have considered interactions as causal analyses, where we seek to identify and estimate the effects of two treatments but also express how the effect of one treatment depends on the other treatment. What must be true for such an analysis to be causally identified? We now use potential outcomes to express the identification conditions. First, Y(d, w) is the potential outcome for Y given the two treatments D and W. Identifiability of the causal interaction above requires the following assumption:<sup>1</sup>

A1. [Unconfoundness] Potential outcomes are independent of both treatments, formally we assume:

$$Y(d, w) \perp D, W$$

A causal interaction captures the joint causal effect of both treatments, and as such, identification requires that the potential outcomes be independent of *both* treatments. Randomized factorial experiments are one way to satisfy this assumption by design. In a factorial experiment, two or more treatments are randomly administered to all study subjects as they were in the Brader, Valentino, and Suhat (2008) study outlined above. Under this type of experiment, this assumption would be satisfied by design. In an observational study, however, this would require an identifica-

<sup>&</sup>lt;sup>1</sup>Our notation implicitly assumes that the potential outcomes are well defined such that we also assume that that the SUTVA holds (Rubin 1986).

tion strategy for each treatment. For example, if identification proceeded under the selection under observables, the investigator would need to articulate the set of covariates needed for identification for both treatments. Alternatively, the researcher would need to outline a different identification strategy for each treatment. For example, one might use an instrumental variables identification strategy for one treatment and selection on observables strategy for the other treatment. The mix of identification strategies isn't important. What is important is that for any study that focuses on causal interaction, there must be an identification strategy for each treatment. In settings with observational data, articulation of multiple credible identification strategies may be difficult.

## 4 Effect Modification

Next, we outline effect modification. Now, we define D as a treatment, and W as baseline covariate measured before D. In a study of effect modification, the analyst is interested in whether the causal effect of D varies with levels of W. Again assuming Y is binary, effect modification would be present if for some w not equal to w' the following holds:

$$P[Y = 1|D = 1, W = w] \neq P[Y = 1|D = 1, W = w']$$

The equation above formalizes the fact that the causal effect of D differs across levels of W. Again, the statistical analysis of effect modification is possible using the following regression model:

$$Y = \beta_0 + \beta_1 D + \beta_2 W + \beta_3 DW.$$

Moreover, the use of the following marginal effect

$$\frac{\partial Y}{\partial D} = \beta_1 + \beta_3 W$$

allows the analysts to easily express how the effect of D varies with W. Next, we consider identifiability for effect modification.

#### 4.1 Identification Conditions

In a study of effect modification, we are only concerned with a single causal effect, in this case the effect of D. To identify the effect of D on Y, requires the following assumption:

A2. [Unconfoundness] Potential outcomes are independent of the treatment, formally we assume:  $Y(d) \perp D$ 

Identification for effect modification only requires that the treatment D be independent of potential outcomes. Even if the relationship between W and Y is confounded, if the above assumption holds, then the effect of D is identified both unconditionally, but also conditional on W.

Unlike a causal interaction, in a study of effect modification, we seek to understand whether a single treatment varies along levels of a baseline covariate. Under this design, the focus is on a single intervention, and the analyst only needs a single identification strategy for treatment D. Here, W is a baseline covariate for which we do not envision potentially intervening on, thus we do not need an identification strategy for the effect of W. Moreover, under effect modification the symmetry of interactions no longer holds. It may hold mathematically, but it does not hold casually. This implies that it is nonsensical under effect modification to declare that the treatment modifies the effect of the baseline covariate. Here, the baseline covariate is clearly the effect modifier of the treatment. While the investigator might calculate the marginal effect for W on Y given D, this is a nonsensical quantity if D is the treatment and W is an effect modifier.

In terms of actual practice, there is another important aspect to studies of effect modification that is critical. Analysts also must ensure that effect modifiers are measured at baseline and thus unaffected by the treatment. It is generally understood that adjustment for post-treatment variables can be an important source of bias (Rosenbaum 1984). This holds true in studies of effect modification as well: if W is post-treatment and possibly effected by D, this may result in bias. Therefore, the effect modifier should be measured temporally prior to the treatment and thus unaffected by the treatment. While it is possible for post-treatment covariates to be effect modifiers, it requires careful attention to additional strong assumptions (Stephens, Keele, and Joffe 2016).

It is our contention that under current practice, investigators generally do not articulate whether the study focuses on causal interaction or effect modification. As we have outlined above, the distinction is critical. First, the distinction matters for the number of identification strategies required. Under causal interaction, multiple identification strategies are needed. Under effect modification, a single identification strategy will suffice. However, care must be taken to ensure that effect modifiers are pre-treatment covariates. Second, the distinction matters for interpretation. Under causal interaction, symmetry holds, and the researcher may wish to examine how causal effects vary across levels of each treatment. Alternatively, investigators may test for the presence of a sufficient cause interaction. If effect modification is the focus of the study, then symmetry does not hold and researchers should only interpret effects of the single treatment across levels of the effect modifier. Next, we review the political science literature to understand whether researchers follow these practices.

#### 5 Review of Current Practice

To get a sense of current practice in the discipline, we performed a brief overview of interaction analyses in the political science literature. Our review is not meant to be exhaustive review of practice in the discipline, but it does serves as a snapshot of current practice in leading journals in the field. For our literature review, we used the same 22 articles from another recent overview of interaction analyses in Hainmueller, Mummolo, and Xu (2016b). The papers were all from leading journals in the field. In our literature review, we first performed a basic classification. That is, we categorized whether the interaction in the analysis paper was one of causal interaction or effect modification. As part of this classification, we also examined whether any of the authors whether authors appeared to be aware of the distinction between causal interactions and effect modification. Then, for each type of analysis, we examined whether the authors paid attention to the requisite details of identification. For causal interaction analyses, this consists of articulating the identification strategy for each treatment that makes up the components of the causal interaction. For effect modification, this includes outlining the identification strategy for the main treatment of interest and ensuring that the effect modifier is measured at baseline and thus un-

affected by the treatment. We think these are the key elements of best practice. We also paid attention to the form of the marginal effects presented. That is, we recorded whether analysts present marginal effects for both treatments when looking at causal interactions. Finally, for effect modification studies, we kept track of whether analysts only present marginal effects with respect to the treatment and not the effect modifier.

In the 22 studies that we reviewed, we classified 17 of the analyses as effect modification, two as studies of causal interactions, and in three cases we found it difficult to classify the results based on the description contained in the article. As we expected, none of the articles frame their analysis using the distinction between causal interactions and effect modification. Neither of the causal interaction articles outlined identification strategies for both treatments. Among the 17 studies of effect modification, we found that analysts did generally treat one covariation as the effect modifier and the other as a treatment. That is, investigators typically only present marginal effects for the treatment and not the effect modifier. What was lacking, however, was a clear discussion of whether the effect modifier was clearly pre-treatment and thus un-affected by the treatment. We only found one article that focused on this issue. Thus, in terms of actual practice, it appears that analysts need to better separate the treatment of interest from effect modifiers in terms of ensuring that those effect modifiers are unaffected by treatment.

# 5.1 Review of Applications

Next, we review two examples from from our literature review to demonstrate the importance of distinguishing between causal interactions and effect modification.

#### 5.1.1 Incumbent Campaign Spending in U.S. House Elections

Kim and LeVeck (2013) explore why spending in House elections has grown dramatically since the early 1970's. Their analysis focuses on three factors: District Partisanship, Uncertainty in Party Reputation, and Incumbent-Party Distance. In their empirical model, they measure district partisanship using presidential vote share, while uncertainty in party reputation is measured using the standard deviation of a party's DW-Nominate scores, and incumbent-party distance is the difference between an incumbents first dimensional DW-Nominate score and their party's

mean DW-Nominate scores in each Congress prior to the election. In their empirical model, they specify three interactions: an interaction between district partial and uncertainty in party reputation, an interaction between incumbent-party distance and district partial partial and interaction between uncertainty in party reputation and incumbent-party distance.

While not explicitly stated, the identification strategy is clearly one of selection on observables as they control for other variables to ensure they have a "fully specified statistical model." However, the exact nature of the identification strategy depends on whether they intend to capture a causal interaction or effect modification. They are silent on this difference, but we infer the following causal structure.

In the analysis, they present three marginal effects plots. In two of these plots, they plot the marginal effect of uncertainty in party reputation and incumbent-party distance as a function of district partisanship. In the third, incumbent-party distance is plotted as function of uncertainty in party reputation. This suggests, they view district partisanship as an effect modifier, with a causal interaction between uncertainty in party reputation and incumbent-party distance.

Let's assume this is what the authors intended. If so, this implies that they need an identification strategy for uncertainty in party reputation, and incumbent-party distance, but not for district partisanship. That is, they need to control for all factors that confound the uncertainty in party reputation and incumbent spending relationship, and all the factors that confound the incumbent-party distance and incumbent spending relationship. If district partisanship is an effect modifier, they only need ensure that it is measured at baseline so that it is not effected by incumbent-party distance or uncertainty in party reputation. However, since it is not a treatment of interest, they need not articulate an identification strategy for the effect of district partisanship on incumbent spending. Again, our point is that understanding the difference between causal interactions and effect modification is critical for specifying the necessary causal identification assumptions.

#### 5.1.2 Returns to Office in China

Truex (2014) considers the question of whether there are returns to office in an authoritarian parliament using data from China's National People's Congress (NPC). Specifically, he focuses on

whether having a chief executive officer (CEO) as a member of the NPC is financially advantageous for Chinese firms. To that end, he compares the financial returns for companies with a CEO that is a member of the NPC and to companies that do not have representation in the NPC. He also conducts an interaction analysis between the NPC membership treatment and state-ownership and revenue.

While he also does not invoke the distinction between a causal interaction and effect modification, it is clear that the interaction analysis performed in one of effect modification. That is, the author presumes that and tests whether the NPC membership treatment effect varies with the level of state-ownership and revenues. Focusing on effect modification greatly simplifies the approach to identification. Here, the analyst need only focus on finding an identification strategy for the NPC, which here is done via a fixed effects design. Both effect modifiers are measured at baseline, and as such there is no need to outline an identification strategy for either one.

## 5.2 Application Summary

These two applications demonstrate the challenge of a causal interaction analysis in observational settings. While opinions vary as to the plausibility of establishing causal effects with observational data, it is generally thought to be a challenging task (Gerber, Green, and Kaplan 2004; Keele 2015; Sekhon 2009). Studies of causal interactions with observational data must develop plausible identification strategies for multiple causal effects. In the first application, the authors must argue that they can control for the set of confounders such that the effect of Uncertainty in Party Reputation on incumbent spending and the effect of Incumbent-Party Distance on incumbent spending are both as-if randomly assigned. In general that seems to be a difficult task. In the second application, the author must develop a plausible identification strategy just for the effect of NPC membership. We do not mean to minimize the difficulty of that task, but, at least, the development and justification of additional identification strategies is unnecessary.

# 6 Discussion and Recommendations for Practice

When analysts present interactions in the context of causal analyses, they should clearly delineate whether a proposed interaction analysis is a causal interaction or instead effect modification.

Critically, the symmetry between variables in the interaction depends on the form of the analysis. Studies of effect modification may appear symmetric, but given identification assumptions are not. While mathematical symmetry holds in studies of effect modification, causal symmetry is nonsensical if one of the variables is meant to modify a causal effect. However, in a causal interaction analysis, such symmetry holds and dependent marginal effects are appropriate for either variable. Given these facts, we recommend the following:

- 1. First, authors should clearly state whether they are interested in a causal interaction, effect modification or both.
- 2. Second, if an interaction analysis is meant to be a study of causal interaction, researchers should clearly specify the number of treatments and the identification strategy for each treatment.
- 3. Third, if an interaction is meant to be a study of effect modification, an identification strategy is only required for the treatment of interest not for the effect modifier. However, analysts must ensure the effect modifier is measured prior to the treatment to avoid bias from conditioning on a post-treatment covariate. This distinction is likely to be more obvious in a randomized experiment, but must be made clear in an observational study.

In general, we would argue that analyses of causal interactions, in most cases, are best restricted to experimental designs. Focusing on causal interactions in experimental settings allows the analyst to establish the identification conditions via the design. However, even in experimental designs, the causal interaction and effect modification distinction remains important. We would argue that many experimental designs should focus on the effect of joint treatments rather than conceptualizing a causal interaction as a case where one treatment modifies the effect of the other. That is, one use for causal interactions is understanding whether causal effects are stronger in the presence of joint treatments. Moreover, given our review of the literature, it would appear that analysis based on causal interactions are somewhat rare and should perhaps see greater use.

# References

- Berry, William D, Matt Golder, and Daniel Milton. 2012. "Improving tests of theories positing interaction." *Journal of Politics* 74(3): 653–671.
- Brader, Ted, Nicholas A. Valentino, and Elizabeth Suhat. 2008. "What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration." *American Journal of Political Science* 52(October): 959–978.
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political analysis* 14(1): 63–82.
- Braumoeller, Bear F. 2004. "Hypothesis testing and multiplicative interaction terms." *International organization* 58(04): 807–820.
- Clark, William, Michael Gilligan, and Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14(Summer): 311–331.
- Franzese, Robert, and Cindy Kam. 2009. Modeling and interpreting interactive hypotheses in regression analysis. University of Michigan Press.
- Gerber, Alan S., Donald P. Green, and Edward H. Kaplan. 2004. "The Illusion of Learning From Observational Research." In *Problems and Methods in the Study of Politics*, ed. Ian Shapiro, Rogers Smith, and Tarek Massoud. New York: Cambridge University Press pp. 251–273.
- Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2016a. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.".
- Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2016b. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.".
- Keele, Luke J. 2015. "The Statistics of Causal Inference: A View From Political Methodology." *Political Analysis* 23(3): 313–335.
- Kim, Henry A, and Brad L LeVeck. 2013. "Money, Reputation, and Incumbency in US House Elections, or Why Marginals Have Become More Expensive." *American Political Science Review* 107(03): 492–504.
- Rosenbaum, Paul R. 1984. "The Consequences of Adjusting For a Concomitant Variable That Has Been Affected By The Treatment." *Journal of The Royal Statistical Society Series A* 147(5): 656–666.
- Rubin, Donald B. 1986. "Which Ifs Have Causal Answers." *Journal of the American Statistical Association* 81(December): 961–962.
- Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." Annual Review of Political Science 12: 487–508.
- Stephens, Alisa, Luke J. Keele, and Marshall Joffe. 2016. "Estimating Post-Treatment Effect Modification With Generalized Structural Mean Models." *Journal of Causal Inference* in press.

- Truex, Rory. 2014. "The returns to office in a "rubber stamp" parliament." American Political Science Review 108(02): 235–251.
- VanderWeele, Tyler. 2015. Explanation in causal inference: methods for mediation and interaction. Oxford University Press.
- VanderWeele, Tyler J, and James M Robins. 2007. "The identification of synergism in the sufficient-component-cause framework." *Epidemiology* 18(3): 329–339.
- VanderWeele, Tyler J, and James M Robins. 2008. "Empirical and counterfactual conditions for sufficient cause interactions." *Biometrika* 95(1): 49–61.