Political Science Research and Methods (2021), **9**, 641–649
doi:10.1017/psrm.2020.12

ORIGINAL ARTICLE

Causal interaction and effect modification: same model, different concepts

Luke Keele[1]* and Randolph T. Stevenson[2]

[1]University of Pennsylvania, Philadelphia, PA 19104, USA and [2]Rice University, Houston, TX 77251, USA
*Corresponding author. Email: luke.keele@gmail.com

(Received 24 October 2019; revised 10 January 2020; accepted 22 January 2020; first published online 21 April 2020)

Abstract

Social scientists use the concept of interactions to study effect dependency. In the causal inference literature, interaction terms may be used in two distinct type of analysis. The first type of analysis focuses on causal interactions, where the analyst is interested in whether two treatments have differing effects when both are administered. The second type of analysis focuses on effect modification, where the analyst investigates whether the effect of a single treatment varies across levels of a baseline covariate. While both forms of interaction analysis are typically conducted using the same type of statistical model, 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 where we show that the form of the interaction is critical to proper interpretation of empirical results.

Keywords: Causal inference; Interaction; effect modification

1. Introduction

Interaction analyses are used in political science to study effect dependency. Currently, there is a large literature that provides guidelines for the estimation and interpretation of interaction analyses based on statistical models (Braumoeller 2004; Brambor et al. 2006; Clark et al. 2006; Franzese and Kam 2009; Berry et al. 2012; Hainmueller et al. 2019). In political science, however, there are few linkages between the literature on interaction analyses and the literature on causal inference. In epidemiology, a large literature has focused on the interplay between interaction analyses and causal inference. In the research on causal inference and interactions, one important distinction is between the concepts of causal interaction and effect modification (VanderWeele and Robins 2007, 2008; VanderWeele 2015). In a causal interaction analysis, analysts are interested in whether the effect of multiple treatments applied simultaneously differ from when these treatments are applied individually. Under effect modification, an investigator focuses on whether the effect of a single treatment varies as a function of baseline covariate or covariates. Critically, while both forms of interactions use identical statistical models, the attendant assumptions and interpretation of statistical results differ in significant ways across the two types of analysis.

In this article, we formally describe the difference between 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 recognize this distinction. We then review two examples from the political science literature to demonstrate how causal

© The European Political Science Association 2020.

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

identification assumptions influence an interaction analysis. We conclude with recommendations for practice.

2. Statistical interactions

The concept of statistical interaction is typically articulated within the context of a linear or generalized linear model as means of operationalizing what we call "effect dependency." Next, we outline the standard framework to define statistical interactions. To that end, we define an outcome Y_i and two covariates D_i and W_i, for a population indexed from $i, \cdots, n$. For simplicity, we take indexing by i as given and drop the i subscript. Next, we define the relationship between these three variables using the following linear model:

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

where ϵ is an error term. Effect dependency can then be expressed by taking the marginal effect with respect to either D or W (Brambor et al. 2006). For example, the following two marginal effects express 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 equation, 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. Consistent with the literature on interactions in regression models, we have thus far used the term "marginal effect" to describe the quantities from (1). However, it is important to note that the term marginal effect is ambiguous. Marginal effect can be used for a descriptive summary of the conditional expectation of Y given D and W. Here, the term describes the partial derivatives with no causal interpretation. Alternatively, marginal effects may describe some pattern of causal influence of D, W, or both on Y. Below, we will focus on when these terms can be given causal interpretations.

Next, it is important to note, that the statistical model in (1) is completely agnostic as to whether the marginal effect of D depends on W or vice versa. That is, the statistical model itself does not provide any insight into whether the effect dependency is symmetric or not. The advice in the political science literature tends to be agnostic about the direction of effect dependency. As Franzese and Kam (2009: 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.

One of the key points that we make below is that this statement is both correct and incorrect depending on the context. However, context is important in two distinct ways. First, it depends on whether an investigator is conducting an analysis that is focused on causal inference. If estimates of causal effects are not the goal of the study then interactions are indeed symmetric. That is, when the marginal effects are conditional expectations with no causal interpretation, this symmetry holds. However, if causal effects are the focus, then symmetry cannot be taken for granted.

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

In particular, 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, the distinction depends on the causal identification assumptions.

3. Causal interactions

First, we review the concept of causal interaction. See VanderWeele (2015) for a complete treatment. Above, when we discussed interactions, we were agnostic as to whether we were interested in causal effects, but now we let D and W denote two binary treatments and the goal is to identify and estimate the causal effects of these two treatments on Y. We also assume that both treatments are administered to the same population of subjects. That is, when either D or W is zero, treatment is absent, but when either is present, one of two of 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 use Y to denote the outcome, and Y_{dw} is the counterfactual outcome where D is set to d and W is set to w. For simplicity, we assume that there are two levels for D and W which we denote as 0 and 1.

The two treatments D and W are said to *interact* if the following contrast holds:

$$E[Y_{1,1}] - E[Y_{0,1}] \neq E[Y_{1,0}] - E[Y_{0,0}]$$

A causal interaction captures the case where the effect of the two treatments together is different from the effect of the two treatments considered separately. In the context of the GOTV application, a causal interaction is present, if the effect of a phone call *and* a postcard is different from the separate effects of a phone call and postcard. Critically, for a causal interaction, the investigator is considering on interventions for *both* D and W. Valid estimates of a causal interaction require certain identification conditions to hold. We outline those conditions in more detail below. However, random assignment of the two treatments is one way to identify a causal interaction.

Next, assuming the identification conditions hold, we can write a causal interaction in terms of observed data. We assume that Y is binary, and we denote the conditional probability for Y as: $p_{\text{dw}} = P[Y = 1 \mid D = d, W = w]$. While Y is binary in this example, the generalization to non-binary Y is straightforward. Following VanderWeele (2015), the causal interaction between D and W on Y is

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

This equation can be rewritten more compactly as

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

This estimate of a causal interaction captures the extent to which the effect of the two treatments together exceeds the effect of each treatment individually. 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, subjects are exposed to an active treatment as is the case in the GOTV example. When this is true, $p_{11} - p_{10} - p_{01} + p_{00}$ 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:

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

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

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. In a causal interaction analysis, interest focuses on the joint effect of the two treatments. If $p_{11} - p_{10} - p_{01} + p_{00}$ 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. In the GOTV example, we might expect the interaction to be super-additive, since the effect of both a phone call and a postcard may 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 for statistical interactions in Equation (1). 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 linear model imposes no functional form assumptions.

A study by Brader et al. (2008) provides 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 about 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 were exposed to a news story that emphasized the costs of immigration. Thus, under this treatment the tone of the news story varied. 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.

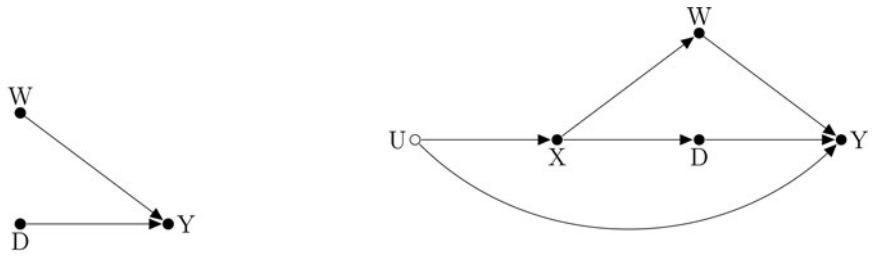
A simple empirical analysis indicates that these two treatments interact. One outcome in the study was a binary measure of whether subjects decided to send their member of Congress an anti-immigration message. In the control condition, 32 percent of subjects selected to send an anti-immigration message. Under the first treatment, where subjects viewed a news story with a negative tone about immigration, the percentage of subjects who send an 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 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. Individually, then, the treatments do not appear to cause participants to send a message to his or her member of Congress.

In the condition where the subjects were exposed to both the tone and racial cue treatments, however, the percentage of people who sent an anti-immigration message increases relative to the control condition by nearly 14 percentage points. As such, while the individual treatments have little effect, the joint treatments exert 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). Here, a causal interaction analysis reveals that there are important features of these two treatments that affect the outcome in a way that is not true when these treatments are administered separately.

3.1. Identification conditions

Next, we present some conditions under which a causal interaction can be identified. Here, we focus on what we think are two prototypical identification conditions for causal interactions. We present identification conditions using causal graphs. See Pearl (2009), Elwert (2013), and

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press



(a) Causal Interaction Identified Case 1 (b) Causal Interaction Identified Case 2

Fig. 1. Two DAGS where causal interactions are identified.

Morgan and Winship (2014) for in-depth reviews of causal graphs. Figure 1 contains two causal graphs. In panel 1(a), we present the most straightforward way in which to identify a causal interaction. In this graph, there are no variables that are confounders between either D or W and Y. How might this scenario arise? This graph would hold if both D and W were randomly assigned. These conditions would hold by design in a randomized trial such as in Brader et al. (2008).

In panel 1(b), we present an example where a causal interaction would be identified without randomization. In this graph, X blocks all the back door paths for both treatments. Therefore, so long as one conditions on X, the total effects of D and W are identified and the causal interaction can be estimated from the data consistently. The critical point, and one that differs from the identification requirements for effect modification described in the next section, is that a causal interaction requires a valid identification strategy for *both* treatments. That is, the analyst must convincingly argue that the total effects for both D and W are identified. This may prove to be challenging outside of a randomized experiment.

4. Effect modification

Next, we focus on effect modification. A wide variety of terms are used in the literature to describe an analysis of effect modification. A sampling of the terms used are "moderation," "conditional average treatment effect," and "heterogenous effects." Now, we define D as a treatment, and W as baseline covariate measured before D is applied. Here, W is often referred to a modifier or "effect modifier." Under effect modification, the focus is on the causal effect of D alone, so we only consider potential outcomes of the following form: Y_d. Effect modification is said to be present when the following inequality holds:

$$E[Y_1 \mid W = w] - E[Y_0 \mid W = w] \neq E[Y_1 \mid W = w'] - E[Y_0 \mid W = w']$$

In a study of effect modification, the analyst is interested in whether the causal effect of D varies across the levels of W. Under effect modification, there is a clear asymmetry between D and W. The analysis is focused on the causal effect of D and the role of W only concerns whether the effect of D varies across the 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

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

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

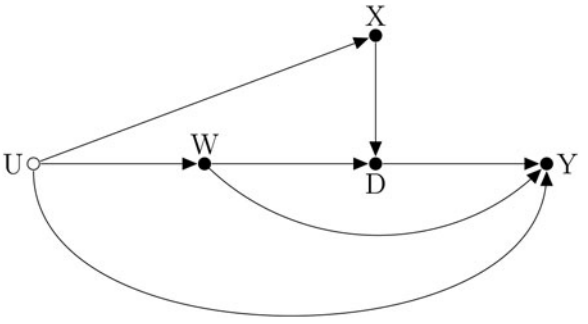


Fig. 2. Causal graph where effect modification is identified but a causal interaction is unidentified.

allows the analysts to easily express how the effect of D varies with W. Next, we consider identifiability conditions 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. As such, identification conditions for effect modification focus on the single treatment of interest. Figure 2 contains a causal graph where the causal effect of D is identified via conditioning but the effect of W is not. That is, if we condition on X and W, the effect of D is identified, but the effect of W remains unidentified due to open backdoor paths between W and Y.

The critical point is that unlike a causal interaction analysis, in a study of effect modification, we seek to understand whether a single treatment varies along the levels of a baseline covariate. Under this design, the focus is on a single intervention, and the analyst only needs an identification strategy for treatment D. Here, W is a baseline covariate that the analyst does not think of as a treatment; instead it is, typically, a background characteristic over which the effect of the treatment may vary. Thus we do not need an identification strategy for the effect of W. Moreover, under effect modification, the symmetry of interactions no longer holds. This symmetry may hold mathematically, but it does not hold casually. That is, it is nonsensical under effect modification to declare that the treatment modifies the effect of the baseline covariate. Here, W is clearly specified as a modifier of the treatment effect of D. 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 critical aspect to the studies of effect modification. Analysts also must ensure that effect modifiers are measured at baseline and thus unaffected by the treatment. It is well understood that adjustment for post-treatment variables can be an important source of bias (Rosenbaum 1984). This holds true in the 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. See Acharya et al. (2016) and Montgomery et al. (2018) for details of this type of bias. While it is possible for post-treatment covariates to be effect modifiers, it requires careful attention to additional strong assumptions (Stephens et al. 2016). Finally, effect modification and causal interaction are not mutually exclusive concepts. One could be primarily interested in causal interaction but test for effect modification by a third baseline covariate.

It is our contention that currently, investigators generally do not articulate whether their interactions should be thought of as causal or effect modification and so often do not address the appropriate causal identification questions. As we have outlined above, however, 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

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

are pre-treatment covariates. Second, the distinction matters for interpretation. Under causal interaction, the researcher focuses on the joint effects of two treatments. If effect modification is the focus of the study, then symmetry does not hold and researchers should only interpret the effects of the single treatment across the levels of the effect modifier. Next, we review the political science literature to understand whether researchers follow these practices.

5. Review of current practice

Next, we performed an overview of interaction analyses in the political science literature. For our review, we used the same 22 articles that were used in another recent overview of interaction analyses Hainmueller et al. (2019). First, we categorized whether the interaction in each paper was meant to capture either causal interaction or effect modification. We also recorded whether the 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 unaffected by the treatment. We also paid attention to any marginal effects that were presented. That is, we recorded whether analysts present marginal effects for both treatments under causal interaction, and whether analysts only present marginal effects with respect to the treatment under effect modification.

In the 22 studies that we reviewed, we classified 17 of the analyses as effect modification, and two as studies of causal interactions. In three cases, we found it difficult to classify the results based on the description contained in the article. None of the articles framed their discussion in terms of the distinction between causal interaction and effect modification. Moreover, neither of the causal interaction articles outlined identification strategies for both treatments. For the effect modification analyses, we found that analysts did generally treat one covariate as the effect modifier and the other as a treatment. Only one article provided a clear discussion of whether the effect modifier was clearly pre-treatment and thus unaffected by the treatment. 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 our literature review to demonstrate the importance of distinguishing between causal interactions and effect modification.

5.1.1. Incumbent campaign spending in US house elections

Kim and LeVeck (2013) explore why spending in House elections has grown dramatically since the early 1970s. Their analysis focuses on three factors: *District Partisanship*, *Uncertainty in Party Reputation*, and *Incumbent-Party Distance*. 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 incumbent's first dimensional DW-Nominate score and their party's mean DW-Nominate scores in each Congress prior to the election. They specified three interactions: an interaction between district partisanship and uncertainty in party reputation, an interaction between incumbent-party distance and district partisanship, and an 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." While the authors do not clarify whether they are focused on effect modification or casual interaction, based on the marginal plots, we inferred that the authors view

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

district partisanship as an effect modifier; with a causal interaction between uncertainty in party reputation and incumbent-party distance.

Let us 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 to ensure that it is measured at baseline so that it is not effected by incumbent-party distance or uncertainty in party reputation. 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 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. He compares the financial returns for companies with a CEO that is a member of the NPC 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 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 to only focus on finding an identification strategy for the NPC. Moreover, both effect modifiers are measured at baseline.

5.2. Application summary

Our review of the literature reveals several important points. First, it appears that authors tend to focus on effect modification over casual interaction. Second, our review demonstrates the challenges of a conducting causal interaction analysis in observational settings. 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

Here, we have delineated the difference between studies of causal interaction and effect modification. We have noted a number of subtle differences between these two concepts. Given these facts, we make a set of recommendations below. It is our contention that the analysis of both experiments and observational studies should follow these guidelines.

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.

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press

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 the analyses of causal interactions, in most cases, are best restricted to experimental designs. However, even in experimental designs, the causal interaction and effect modification distinction remain important. Moreover, given our review of the literature, it would appear that the analyses based on causal interactions are somewhat rare and should perhaps see greater use.

Acknowledgments. We would like to thank Matt Golder for comments on an earlier version of this manuscript.

References

Acharya A, Blackwell M and Sen M (2016) Explaining causal findings without bias: detecting and assessing direct effects. *American Political Science Review* **110**, 512–529.

Berry WD, Golder M and Milton D (2012) Improving tests of theories positing interaction. *Journal of Politics* **74**, 653–671.

Brader T, Valentino NA and Suhat E (2008) What triggers public opposition to immigration? Anxiety, group cues, and immigration. *American Journal of Political Science* **52**, 959–978.

Brambor T, Clark WR and Golder M (2006) Understanding interaction models: improving empirical analyses. *Political Analysis* **14**, 63–82.

Braumoeller BF (2004) Hypothesis testing and multiplicative interaction terms. *International Organization* **58**, 807–820.

Clark W, Gilligan M and Golder M (2006) A simple multivariate test for asymmetric hypotheses. *Political Analysis* **14**, 311–331.

Elwert F. (2013) Graphical causal models. In Stephen L. Morgan (ed.). *Handbook of Causal Analysis for Social Research*. Amsterdam: Springer, pp. 245–273.

Franzese R and Kam C (2009) *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor, MI: University of Michigan Press.

Hainmueller J, Mummolo J and Xu Y (2019) How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis* **27**, 163–192.

Kim HA and LeVeck BL (2013) Money, reputation, and incumbency in US house elections, or why marginals have become more expensive. *American Political Science Review* **107**, 492–504.

Montgomery JM, Nyhan B and Torres M (2018) How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science* **62**, 760–775.

Morgan SL and Winship C (2014) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd Ed. New York, NY: Cambridge University Press.

Pearl J (2009) Causal inference in statistics: an overview. *Statistics Surveys* **3**, 96–146.

Rosenbaum PR (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**, 656–666.

Stephens A, Keele LJ and Joffe M (2016) Estimating post-treatment effect modification with generalized structural mean models. *Journal of Causal Inference*. in press, unpublished Manuscript.

Truex R (2014) The returns to office in a "rubber stamp" parliament. *American Political Science Review* **108**, 235–251.

VanderWeele T (2015) *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press.

VanderWeele TJ and Robins JM (2007) The identification of synergism in the sufficient-component-cause framework. *Epidemiology* **18**, 329–339.

VanderWeele TJ and Robins JM (2008) Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika* **95**, 49–61.

Cite this article: Keele L, Stevenson RT (2021). Causal interaction and effect modification: same model, different concepts. *Political Science Research and Methods* **9**, 641–649. https://doi.org/10.1017/psrm.2020.12

https://doi.org/10.1017/psrm.2020.12 Published online by Cambridge University Press