social sciences

Article

Interactions in Generalized Linear Models: Theoretical Issues and an Application to Personal Vote-Earning Attributes

Tsung-han Tsai [1] **and Jeff Gill** [2],*

[1] Department of Political Science, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899, USA; E-Mail: t.tsai@wustl.edu

[2] Department of Political Science, Division of Biostatistics, and Department of Surgery, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899, USA

* Author to whom correspondence should be addressed; E-Mail: jgill@wustl.edu; Tel.: +1-314-935-9012; Fax: +1-314-935-5856.

Abstract: There is some confusion in political science, and the social sciences in general, about the meaning and interpretation of interaction effects in models with non-interval, non-normal outcome variables. Often these terms are casually thrown into a model specification without observing that their presence fundamentally changes the interpretation of the resulting coefficients. This article explains the conditional nature of reported coefficients in models with interactions, defining the necessarily different interpretation required by generalized linear models. Methodological issues are illustrated with an application to voter information structured by electoral systems and resulting legislative behavior and democratic representation in comparative politics.

Keywords: generalized linear models; interaction effects; first differences; personal vote-earning attributes

1. Introduction

Generalized linear models, limited dependent variable models in particular, are now quite common in empirical studies in the social sciences. Multiplicative interaction terms have been routinely incorporated in linear model applications, and are well-understood [1]. Recently, there has been increasing attention paid to the role of these interaction terms in such nonlinear forms in social science literatures, including

Berry *et al.* [2], Berry and Berry [3], Berry [4], Frant [5], King [6], Brambor *et al.* [7], and Kam and Franzese [8]. However, only a few authors have noted the additional complexity imposed by the conditional nature of interaction specifications in the course of their analyses [9–12]. A central issue is that the interpretation of interactions is more complex in generalized linear models than in the basic linear construct, and it is not widely known that interaction effects are necessarily introduced into all generalized linear models by the link function.

Although interaction effects are inherently imposed through the link function, researchers usually include an interaction term into generalized linear models for theoretical reasons since parameter-based interactions provide interactive effects that are different from the automatic variety introduced by the link function. The presence of both the link function and the interaction terms complicate the interpretation of the relationship between the outcome variable and the explanatory variables. This article reviews the meaning and interpretation of interactions in linear and generalized linear models focusing on their conditional nature. The associated problem of determining statistical significance is discussed and illustrated with examples from an issue centered on interpreting interactions in cross-sectional data, and from a reanalysis of personal vote-earning attributes evaluated with hierarchical interaction specifications.

Failure to account for these automatically generated interactions and their conditional nature can easily produce models that are misleading and wrong: the *individual* printed coefficients and standard errors do not alone convey the full contribution and reliability of a given explanatory variable. Clark *et al.* [13] survey widely distributed journals in one field and find errors to be common. In this article, we suggest reporting interaction effects with first differences for generalized linear models to avoid a misinterpretation.

2. Interactions in Linear Models

The typical treatment of interactions in linear models is to consider the interaction as a product term of the main effect variables. [1] This takes the following form where β_i is the sample-derived estimate of the unknown population parameter $\mathbb{B}_i$:

$$E[Y_i|\mathbf{X}] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} \tag{1}$$

and β_3 is the coefficient estimate corresponding to the product. The complete product term, $\beta_3 X_{i1} X_{i2}$, is called a *first-order interaction* or sometimes a *two-factor interaction*, where for obvious reasons the order is one less than the number of factors. Subject to mild assumptions [16], the sampling distribution of β_3 over its standard error is Student's-t with $N - k - 1$ degrees of freedom. There is no requirement

---

[1] Unfortunately, the term "main effect" is not very revealing or accurate. The coefficient for a non-multiplicative explanatory variable specification given by a variable that is specified as interacting with others, is the contribution for that variable at zero levels of all of the interacting variables. There is no mathematically implied primacy and a more accurate term would be "non-multiplicative" or "single contribution." However, it seems unwarranted to try and change standard vernacular used by an overwhelming number of authors, dating from as far back as Fisher and Mackenzie [14] and Finney [15]. So this term will be used here.

for the form of this interaction term to be a product of the main effects and others have been suggested, such as $\beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2}$ [17] and $\beta_j X_{i1}/X_{i2}$ [18]. [2]

The meaning of interactions in the linear model is actually easier to interpret if Equation (1) is rearranged as follows:

$$E[Y_i|\mathbf{X}] = \beta_0 + \beta_1 X_{i1} + (\beta_2 + \beta_3 X_{i1})X_{i2} \tag{2}$$

If one is interested in the consequence from changes in the explanatory variable X_{i2} on the outcome variable, it is necessary to take the first derivative of Equation (2) with respect to this variable in order to obtain the *marginal* effect as a composite coefficient estimate:

$$\frac{\partial}{\partial X_{i2}}E[Y_i|\mathbf{X}] = \beta_2 + \beta_3 X_{i1} \tag{3}$$

This is useful because it demonstrates that the effect of levels of X_{i2} on the outcome variable is intrinsically tied to specific levels of X_{i1}: the marginal contribution of X_{i2} is conditional on X_{i1}.

Two scenarios can occur, the first when high levels of one variable have an accelerating effect on the other (β_3 has the same sign as β_2), and the second when high levels of one variable have a dampening effect on the other (β_3 has the opposite sign of β_2). So the sign on first-order interaction effects tells us quite a bit about the *conditional* effect that a given explanatory variable has on the outcome variable. The interpretation of a given coefficient's effect is now complicated by the requirement that it occurs at a specified level of the other explanatory variable. Often there are theoretically important levels of X_{i1} that can be substituted into Equation (3). In the absence of some theory-driven level or point of particular interest, the quantiles of the interaction explanatory variable provide convenient points of analysis. [3]

It is important to remember that including an interaction component in linear models (as well as those discussed below) but omitting the corresponding marginal main effects produces model results that are difficult to interpret (see Nelder [23] for a strong critique). This is true even if a main effect is not statistically significant, and even if the researcher has a prior hypothesis that the main effect is not relevant [24]. In the context of Equation (1), such a specification would be $E[Y_i|\mathbf{X}] = \beta_0 + \beta_3 X_{i1}X_{i2}$. This form is equivalent to the assertion that neither of the two explanatory variables has an influence on the outcome variable in the absence of the other [25]. If this is really the case, then essentially $X_{i1}X_{i2}$ is a single variable, not an interaction term between two variables. [4] Unfortunately, omitting lower-order

---

[2] These models use algebraic forms where components such as exponents are estimated (and hence are different from standard polynomial regression, e.g., Fox ([19], S. 14.2)) in order to make local approximations to continuous, multidimensional response surfaces. For extended discussions, see Box and Draper [20], Khuri and Cornell [21], and Cornell and Montgomery [22].

[3] It is also interesting to note that when the multiplicative interaction term $X_{i1}X_{i2}$ is highly correlated with one of the main effects, it is an indication that the other main effect is not varying much (consider the extreme case where one of the "variables" was the constant!).

[4] Sometimes this is exactly the case and perfectly appropriate. For instance, including gross domestic product (GDP) without separate terms for price and quantity in a model specification implies that price and quantity are important only when considered together. A reader would be quite confused to see price × quantity in this situation as if the constituent parts of GDP had some significance beyond their joint contribution but were not worthy of consideration as individual explanatory variables.

terms means that common transformations, such as centering and standardization, actually substantively change the coefficient values: violating the principle of invariance [1,18]. In addition, the variance of such an interaction term is interpretable only in the context of the variance of its (now absent) main effect components [26].

3. Interactions in Generalized Linear Models

The statistical treatment of interaction effects in generalized linear models, dichotomous choice models in particular, is a well-developed tool in bioassay and epidemiology, where researchers are often concerned with the effects of combinations of various behaviors and exposures on disease rates. A typical example is the apparent interaction of exposure to some airborne industrial toxins with smoking as a contributor to lung cancer risk [27]. In this context an interaction is defined as the statistical departure from an additive specification that affects the success rate [28–30]. Here, the incidence of lung cancer is accelerated by the *combination* of smoking and exposure in a non-linear fashion that cannot be adequately modeled by the additive marginal effects of smoking and exposure alone. This characterization is also typical of the type of effects that social scientists are seeking to model, interactions such as: democracy, economic growth, and trade with time [31], party identification with sociotropic evaluation [32], ideology with House committee membership [33], ethnicity with participation [34], and democratic government with war initiation [35].

3.1. The Effect of the Link Function

The bulk of the applications in this area involve dichotomous or polychotomous explanatory factors and their corresponding interaction terms [36]. However, the methodological issue is actually much broader, encompassing outcome variables that are: dichotomous choices, event counts, those resulting from truncated sample spaces, and those defined on bounded subspaces. The family of generalized linear models encompasses most of these forms, and the approach to interactions in generalized linear models provided here gives a uniform treatment for interactions covering nearly all parametric regression-type models specified in law and political science.

Interaction effects are more complicated in generalized linear models due to the link function between the systematic component and the outcome variable. In generalized linear models, the systematic component is related to the mean of the outcome variable by a smooth, invertible function, $g(\cdot)$, according to:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} \qquad \text{where: } \boldsymbol{\mu} = E(\mathbf{Y}|\mathbf{X}) = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{4}$$

This is a very flexible arrangement and it allows the modeling of non-normal, bounded, and non-continuous outcome variables in a manner that generalizes the standard linear model [37,38]. Using the link function, it is possible to change Equation (2) to the more general form:

$$E[Y_i|\mathbf{X}] = g^{-1}(\beta_0 + \beta_1 X_{i1} + (\beta_2 + \beta_3 X_{i1})X_{i2}) \tag{5}$$

which obviously reduces to the linear model if $g^{-1}(\cdot)$ is the identity function. Common forms of the link function for different assumed distributions of the outcome variable are $g(\boldsymbol{\mu}) = \log(\boldsymbol{\mu})$ for Poisson

treatment of counts, $g(\boldsymbol{\mu}) = -\frac{1}{\mu}$ for modeling truncation at zero with the gamma, and logit ($\log(\frac{\mu}{1-\mu})$), probit ($\Phi^{-1}(\boldsymbol{\mu})$), or cloglog ($\log(-\log(1 - \boldsymbol{\mu}))$) for dichotomous forms.

A less well-understood ramification of interactions in generalized linear models is that by including a link function, *the model automatically specifies interactions* on the natural scale of the linear predictor (though not necessarily on the transformed scale of the linear predictor). To see that this is true, revisit the calculation of the marginal effect of a single coefficient by taking the derivative of Equation (5) with regard to some variable of interest as done for the linear model (3) but without an explicitly specified multiplicative term for the interaction. If the form of the model implied no interactions, then this calculation would produce a marginal effect free of other variables, but this is clearly not so:

$$\frac{\partial}{\partial X_{i2}} E[Y_i|\mathbf{X}] = \frac{\partial}{\partial X_{i2}} g^{-1}(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}) = (g^{-1})'(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})\,\beta_2 \qquad (6)$$

This demonstrates that the presence of a link function requires the use of the chain rule and therefore retains other terms on the right-hand side besides β_2. If the link function were the identity function, then its derivative with respect to X_{i2} would in Equation (6) simplify to β_2, as in the standard linear model without specified interaction components. However, with the generalized linear model there are always partial effects for a given variable that are dependent on the levels of the other explanatory variables.

3.2. Illustration of the Link Function Effect

To give a more specific example that illustrates the effect of the link function, consider a dichotomous outcome variable indicating personal vote-earning attributes (PVEA) of legislators—their birthplaces and prior local electoral experience [39]. It is argued that electoral rules affect candidates' strategy to rely on either personal or party reputation in a campaign [40]. Specifically, legislators in systems encouraging intraparty competition are likely to cultivate their personal votes, whether through attributes or legislative behavior [39,41].

Consider proportional representation (PR) systems where two key features are related to the degree of intraparty competition. The first is the type of list that parties present: *closed* and *open lists*. Since intraparty competition in open lists is more intense than that in closed ones, politicians are more likely to engage in personal vote-seeking behavior in open lists. Therefore, the probability that legislators and legislative candidates exhibit PVEA in open lists is higher than that in closed lists. The second feature is the *district magnitude* (e.g., small, medium, and large). For the purpose of illustration, assume that the effect of magnitude on the outcome variable is positive in both open and closed lists.[5]

---

[5] Notice that, according to Carey [41] and Shugart *et al.* [39], list type and district magnitude interact to affect the incentive of the politicians to cultivate a personal vote and this interaction effect makes open lists and close lists influence PVEA in opposite directions. The hypothesized interaction effect between list type and district magnitude is different from the one introduced by the link function. Specifically, the interaction effect introduced through the link function only changes the *magnitude* of the main effect as can be seen below in Figure 1. However, adding an interaction term might change not only the magnitude but also the *direction* of the main effect. This is because the effect of the interaction term overwhelms the main effect as we show in Section 6. Therefore, the model presented in Equation (7) is not used to test the theoretical argument but only for the purpose of illustration.

A simple probit model with no directly specified interaction term is written as follows:

$$\Pr(\texttt{native} = 1) = \Phi(\beta_0 + \beta_1 \log \texttt{M} + \beta_2 \texttt{open}) \tag{7}$$

where Φ denotes the standard normal cumulative distribution function. This model is depicted in Figure 1, where open and closed lists are separated. If this were the standard linear model, or if the given systematic effect were plotted on the linear systematic scale, then the difference between open and closed lists would be that seen in the left panel of Figure 1, which shows no interactions. However, even though there was not a deliberate specification of an interaction term, it is clear from the right panel of Figure 1 that list type interacts with district magnitude in affecting the probability that legislators and legislative candidates exhibit PVEA on the probit (probability) metric. When the magnitude is at the extremes of the district magnitude scale, open and closed lists are not substantially different from each other in probability. However, for systems with modest levels, the probability of exhibiting PVEA for legislators in open lists is significantly higher than that in closed ones: a difference of about 40%.

Figure 1. Probability models for personal vote-earning attributes. The left panel displays a standard linear model and the right panel displays a probit model.
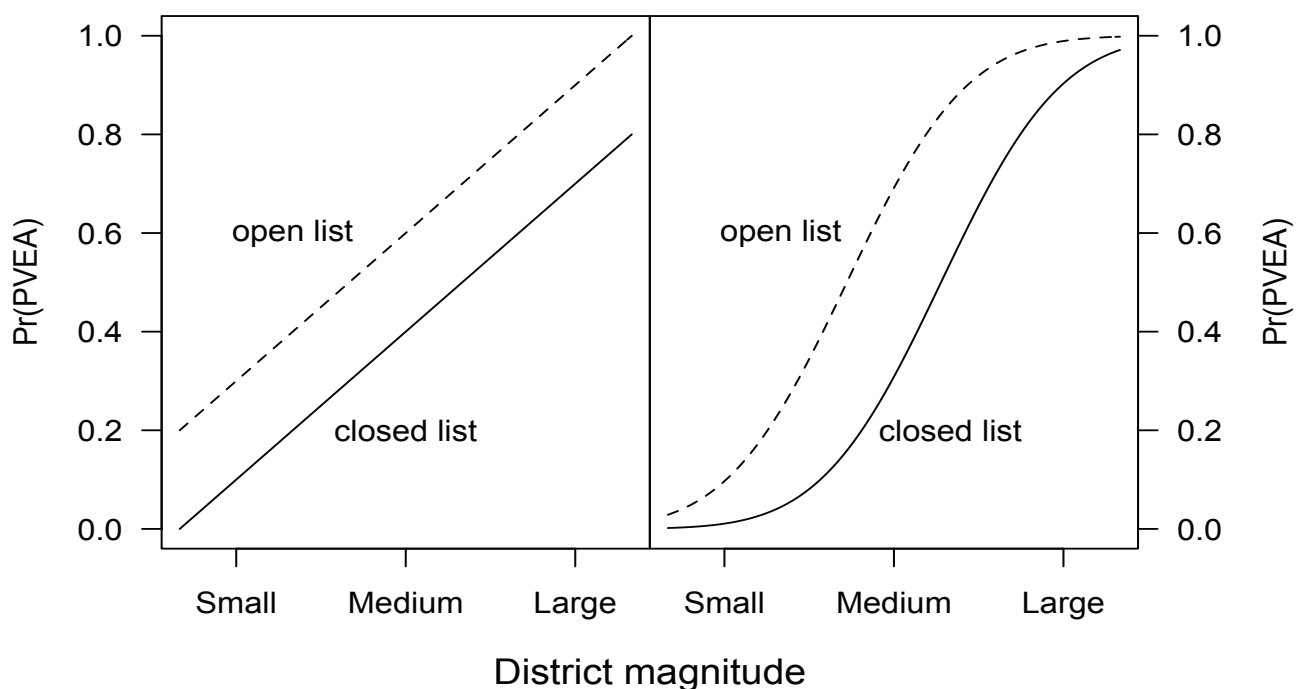
3.3. Interpreting Interaction Effects in Generalized Linear Models

From the preceding discussion it is clear that interactions are naturally produced in generalized linear models, regardless of whether they are recognized or desired. Yet this observation does not really help in *testing* for the existence and statistical reliability of hypothesized interactions, or in determining overall model quality in the acknowledged presence of such terms.

There are two ways of thinking about the outcome variable of interest in generalized linear models. First, we can directly model the expected value of the outcome variable by a link function such as $\mu = g^{-1}(\mathbf{X}\boldsymbol{\beta})$. In the context of generalized linear models, the interpretation of the main and interaction

effects is not straightforward because of the presence of the link function as shown in Equations (5) and (6). Second, since the link function is defined to be smooth and invertible, it is always possible to pass this function back to the left-hand side of Equation (4) and consider a specification in which a function of the expected value of the outcome variable is modeled linearly: $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$. The important difference is that with this form, the automatic interaction terms cancel out on this metric. *This does not mean that such terms do not exist and it does not mean that by strictly considering a model of this form that one does not have to worry about potential interactions* [2]. Instead, it means that if one is interested in a form of the outcome variable that is modified directly by the appropriate link function, and if the desired specification excludes interaction terms, then the model can be greatly simplified.

To more concretely make this point, consider a case where the conceptual outcome variable of interest is dichotomous and the specification uses a logit, probit, or cloglog function as a convenience to link the additive component with support over the entire real line to the necessarily bounded probability of getting a success. The logit specification in this context produces:

$$P(Y_i = 1|\mathbf{X}) = (1 + \exp\left[-\beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2}\right])^{-1} \tag{8}$$

If a model like Equation (8) is calculated with the standard numerical procedure for GLMs (maximum likelihood estimation through iterative weighted least squares, see Fahrmeir and Tutz ([37], p. 42), Green [42], or del Pino [43] for details), then such interactions are always present. Berry *et al.* [2] call such effects *compression* because they push the resulting probability further towards zero or one (depending on the sign). These inadvertent interaction effects are not parameterized in the model, so their influence on this probability does not come with a coefficient that can be observed and tested.

Sometimes interest is centered instead on an indirectly obtained unbounded latent variable. Return to the previous example where the real interest might not lie in the binary outcome of exhibiting PVEA for some particular combination of levels of the explanatory variables, but rather in the effects of particular explanatory variables on the *inclination* to exhibiting PVEA expressed through the relative odds of vote-switching versus not exhibiting PVEA. In logit models this is the log-odds of a successful outcome (*i.e.*, the outcome coded 1), and is obtained by moving the GLM link function to the left-hand side of Equation (8):

$$\log\left[\frac{P(Y_i = 1|\mathbf{X})}{P(Y_i = 0|\mathbf{X})}\right] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \tag{9}$$

This treatment linearly models an unbounded latent variable rather than $P(Y_i = 1|\mathbf{X})$, which is obviously bounded by zero and one. While we produced this specification from algebraic manipulation of Equation (8), it is important to understand that this is really a different model with a different outcome variable altogether [2]. If interest is truly centered on the log-odds, then Equation (9) can be estimated simply with ordinary least squares and no interaction effects are produced from a link function. However, the researcher is making the strong statement that there are no interactions of consequence to be modeled, and therefore no associated hypotheses to be tested. Regarding this specification, McCullagh and Nelder ([38], p. 110) unambiguously state that: "It is important here that x_1 be held fixed and not be permitted to vary as a consequence of the change in x_2." In other words: no interactions.

The specification in Equation (9) does *not* preclude the inclusion of interaction terms. Quite the contrary. Now that the right-hand side is a simple additive form, specifying interaction terms is as straightforward as with the standard linear model discussed in the first section:

$$\log\left[\frac{P(Y_i = 1|\mathbf{X})}{P(Y_i = 0|\mathbf{X})}\right] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} \tag{10}$$

Here there will *only* be the researcher-specified interaction effect $\beta_3 X_{i1} X_{i2}$.

Cox, in particular, advocates the log-odds expression for dichotomous choice specifications with specified interaction terms as an interpretational simplification of the analysis by using an unconstrained outcome variable, "...since such representations have the greatest scope for extrapolation" ([24], p. 21). Nagler [9] calls model components such as $\beta_3 X_{i1} X_{i2}$ in Equation (10) "variable-specific" interaction terms to distinguish them from the automatic variety. If specifying these variable-specific terms in the model leads to improved fit (comparable with likelihood ratio tests), then the model has successfully captured through parameterization at least some of the interaction between variables.

Unfortunately, differing forms of the link function (those that produce other generalized linear models) do not share the naturally intuitive appeal of the log-odds of success outcome variable that the logit model provides. Therefore, researchers who move the link function to the left-hand side of the GLM specification equation purely to avoid considering imposed interaction considerations do so at a great interpretational expense.

3.4. Reporting GLM Interaction Effects with First Differences

The best way to understand generalized linear model results (including those with dichotomous choice outcome variables) is by *first differences* [44], a term that originates in time-series work. The principle of first differences, sometimes called attributable risk, is to select two levels of interest for a given explanatory variable and calculate the difference in impact on the outcome variable holding all of the other variables constant at some value, usually the mean.

This principle is only slightly more complicated with interaction specifications. Since the total effect of any given variable is a composite of its main effect and associated interaction effects, it is necessary to include all interacting coefficients in the first difference calculation, making sure that all other variables in these terms are set at their means in both their main effects and in the interaction effects. Thus, when looking at a variable of interest in a table of first differences, the observed difference includes the main effect *as well as all of the interaction effects that include that particular variable*.

The mechanics of this process are relatively simple. Recall the generalized linear model form in Equation (5): $E[Y_i|\mathbf{X}] = g^{-1}(\beta_0 + \beta_1 X_{i1} + (\beta_2 + \beta_3 X_{i1})X_{i2})$ where the explanatory variable X_2 is of primary interest. The coefficients $\beta_0, \beta_1, \beta_2$ are estimated and X_1 is set at its mean: $\bar{X}_1$. Now select two values of X_2 of interest: $X_2^{[1]}$ and $X_2^{[2]}$. This is usually very straightforward for categorical variables, and only slightly less so for interval measured variables of which one can pick extreme values or quantiles of interest. Of course there are often theoretically important values to assign as well. These assignments allow us to produce two theoretical expected values of the outcome variable:

$$E(Y|\mathbf{X})^{[1]} = g^{-1}\Big(\beta_0 + \beta_1 \bar{X}_1 + (\beta_2 + \beta_3 \bar{X}_1)X_2^{[1]}\Big)$$

$$E(Y|\mathbf{X})^{[2]} = g^{-1}\Big(\beta_0 + \beta_1 \bar{X}_1 + (\beta_2 + \beta_3 \bar{X}_1)X_2^{[2]}\Big) \qquad (11)$$

and take the difference: $FD^{[2,1]} = E(Y|\mathbf{X})^{[2]} - E(Y|\mathbf{X})^{[1]}$. This value therefore gives the expected difference for the outcome variable across the selected range of one explanatory variable of interest, holding the other(s) constant at the mean.

These are really "conditional first differences" here because not only are the other variables held constant at their mean, but also the first difference values and the standard errors of these first differences are calculated by including the effects of the relevant interaction terms. For this example, with X_2, the first difference is calculated by summing the contribution from the main effect as well as the first-order interaction with X_1. This means that the first difference for the variable of interest is conditional on the set value for the other variable that it interacts with. The consequence of this conditionality is that the measure of uncertainty for the first difference is a composite term conditional on more than one singular standard error. [6]

4. A Methodological Controversy in Political Science

The treatment of interaction terms in generalized linear models is at the core of a discussion between Frant [5] and Berry and Berry [3] concerning the specification of dichotomous models with a logit link function applied to state-level lottery data. Berry and Berry assert that deliberately specifying multiplicative interaction components is appropriate for testing hypotheses when the left-hand side specification is a log-odds form such as Equation (9), but when an expected probability is modeled, the automatically generated interaction effects in the marginal first differences for individual coefficients are sufficient for such tests. Frant disagrees, stating that no explicit hypothesis of interaction can be tested by such indirect means in cases where the outcome variable is treated as the expected probability of success. Both Berry and Berry and Frant do agree, however, that when the hypothesis test is for a log-odds outcome variable, any desired test of interaction must be directly specified with a multiplicative term. So their disagreement is really over whether it is necessary to specify the coefficient β_3 in Equation (10), or whether looking at the main effect coefficients alone through first differences is sufficient.

Berry [4] revisits this question with an updated argument for expected probability models. He asserts that specifically including the $\beta_3 X_{i1} X_{i2}$ term *is* necessary and that the previously recommended approach of first differences works only if all terms that specify the variable of interest, including interactions, are included in the calculation. The debate between Berry and Frant is therefore over whether sufficient information to perform a hypothesis test of interactions can be found in the magnitude of the interaction term alone (Frant) or whether it is necessary to use this term and the appropriate main effects terms to judge the impact of the variable of interest at different levels of the interacting variable through first differences (Berry).

---

[6] Since the complete effect of the explanatory variables specified as interacting in the model is conditional on the values of the other co-interacting specified variables, the associated standard error is as well. The derivation of the conditional standard errors for the simple case when there are only two explanatory variables is provided by Friedrich ([1], p. 810) and Jaccard *et al.* ([45], p. 27), and an abstract form is given in various textbooks, e.g., Timm [46].

Berry is correct here. The existence of the link function means that the effect of any particular explanatory variable on the left-hand side probability depends on the values of every other explanatory variable and therefore must be included in a hypothesis test. To see how wrong the singular approach can be, imagine that the main effect coefficients for two explanatory variables are both large and positive: $\beta_1 + + +, \beta_2 + + +$, whereas the interaction coefficient is small and negative: $\beta_3 -$. The singular approach would be to interpret a purely negative (dampening) effect of X_1 on X_2, and vice-versa. However, it is likely that these main effect contributions will overwhelm this term, and *the researcher would have no way of knowing without calculating the full contribution of all interacting variables expressed through the link function using first differences* as done in Equation (11) (Berry finds such a case). This is due to both the magnitude of the coefficients (assumed large here) and the measurement of the variables involved.

Furthermore, some link functions provide directional results that are surprising given the signs on the tabulated coefficients, such as the gamma link function, which imposes a minus sign and an inverse on the linear predictor. In other words, there is no way of fully appreciating the effect of interacting variables until the complete levels (means and first difference values) are wound through the GLM link function using every associated term. There are two prominent mistakes often made here: assuming that only the interaction coefficient is necessary to perform hypothesis tests for interactions (Frant), and neglecting to specify an interaction coefficient at all, assuming that the interaction effects would be revealed through the covariance imposed by the link function (Berry and Berry).

This debate highlights the different ways that Equations (8) and (9) are treated, even though they are *exactly* the same mathematical model of the data. The primary difference lies in how one chooses to think about the outcome variable, and all of these authors agree that the log-odds metric is easier to handle from a modeling perspective, given the caveat that any desired interactions must be specifically stipulated as in the linear form. However, while first differences of log-odds are less intuitive than first differences of probabilities, the ease with which interaction terms can be added to the linear predictor of log-odds specifications makes this form attractive. As long as it is understood that the GLM link function can be made to operate on either side of the equality and that interaction implications differ, there is no added complexity in using both forms.

5. Higher-Order Interactions

There are no mathematical or statistical restrictions to developing higher-order interactions in linear or nonlinear models [18,22,47]. These are coefficients on interaction effects between lower-order interaction effects and other effects as well as interaction effects between interaction effects. Specifying such terms naturally introduces a hierarchy into the model, not just because higher-order interactions are not defined in the absence of the main effects and the lower-order interaction effects that constitute their foundations, but also because the interpretation of these effects depends on the assumed levels of constituent terms. The primary impediment to the widespread use of interaction hierarchies in model specifications is that there is currently no general method for analytically calculating the correct measure of uncertainty for the resulting explanatory effects, although they can sometimes be obtained by

careful simulation [48]. Instead, the analyst is usually required to perform lengthy derivations for each differing specification.

Several motivations exist for developing models with higher-order interactions. They are superior to general polynomial models (of which they are a subclass) because the higher-order terms reveal more about the structure of dependency effects in the explanatory variables [22]. De Leeuw [49] points out that a hierarchical linear model "links the analysis of discrete and continuous variables convincingly, and makes it possible to discuss the interpretation of a great variety of statistical techniques in terms of conditional independence." Constructing hypothesis tests for higher-order interaction terms is no more complicated than for first-order interaction terms [50]. Also, Aiken and West [51] argue that the hierarchical nature of increasing levels of interaction effects easily facilitates exploratory hypotheses as these levels can be increased systematically.

There is an extensive literature on testing the existence of first- and higher-order interactions in factorial experiments and with contingency tables. This is because whenever a factor is crossed with another, three sources of variation are generated: variation from each of the two sources and variation from the residuals after removing the marginal effects [52]. There are well-developed tests for the existence of these interaction effects under a broad range of data types, dimensions, and restrictions [38,53–56], and some authors in this area have paid particular attention to higher-order interactions [50,57]. However, there is considerably less to draw on in terms of generalized linear models with continuous random variables (perhaps the more typical application in social science work). Ironically, these are more difficult cases to analyze because one loses the naturally intuitive appeal of the effect of being in one of a finite number of cells in a multidimensional rectangle of states.

Suppose for the moment that there exist three main effects of interest, and there is a reason to specify a second-order interaction hierarchical model. Using the logit specification of the generalized linear model from Equation (5) produces the following specification for the log-odds of success:

$$\log\left[\frac{P(Y_i = 1|\mathbf{X})}{P(Y_i = 0|\mathbf{X})}\right] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$
$$+ \beta_4 X_{i1} X_{i2} + \beta_5 X_{i2} X_{i3} + \beta_6 X_{i1} X_{i3} + \beta_7 X_{i1} X_{i2} X_{i3} \qquad (12)$$

where the log-odds form is used to make the imposition of multiple interactions more straightforward.

So now in addition to having three first-order interaction effects, there is a second-order interaction effect with contributions from all three explanatory variables. Computationally there is no difficulty specifying and estimating models with higher-order interaction effects such as Equation (12). The interpretation of the second-order interaction effect is straightforward: if one is primarily interested in the explanatory variable X_3, then it is appropriate to consider the coefficient β_7 as the increment to β_3 for a particular combination of values of X_1 and X_2. Actually this term can be interpreted as a first-order interaction between *any* of the first-order interaction effects and the remaining main effect:

$$\beta_7(X_{i1} X_{i2})X_{i3} = \beta_7(X_{i2} X_{i3})X_{i1} = \beta_7(X_{i1} X_{i3})X_{i2} \qquad (13)$$

This seems like a very simple point, but there is more here than just a trivial algebraic rearrangement of parentheses: each of the three explanatory variables in Equation (13) can be individually considered as the primary variable of interest and the other two as "moderator" variables [45].

An interpretational motivation for considering higher-order interaction effects, such as in Equation (13), is that a combination of levels of two variables expressed through a first-order interaction term can alter how a third variable affects the outcome variable. For instance, in the running example, if the interaction effect between list type and district magnitude altered the effect of rookies on the probability of exhibiting personal vote-earning attributes, then it might be important to add this effect to a proposed model specification [39]. As Shugart *et al.* argue, under open lists, it should be more important for rookie legislators to have PVEA to advertise in order to compete with veterans as magnitude increases. Under closed lists, rookies would not require PVEA in high-magnitude districts while they are likely to have those attributes at lower magnitudes. Thus, there exist differing effects by the magnitudes of district, list types, and rookie status, which can be separated out by a second-order interaction effect parameterization. We specifically test these hypotheses in the next section.

It is easy to see that a reasonable number of main effects typical of a social science empirical model can produce a huge number of interaction hierarchies. In fact, the number of possible interactions for $\nu > 1$ main effects is determined by the rapidly increasing function: $\sum_{i=2}^{\nu} \binom{\nu}{i} = 2^{\nu} - \nu - 1$ (applying the binomial theorem). Therefore, the selection of a useful subset of possible interaction effects is a vital part of the art of model specification.

Although it is argued here that interaction effects provide useful information and lead to better empirical models, researchers should use the same sort of theory-driven procedure that is involved in picking main effects. A "kitchen sink" (or, worse yet, some variant of a stepwise) interaction specification approach is likely to lead to a massively parameterized and substantively useless model [22,57].

In addition, just as specifying first-order interaction effects while omitting the corresponding main effects produces poor models, specifying second- and higher-order effects without the component lower-order interaction effects is also ill-advised unless there is a strong theoretical justification [18,58].

6. An Application to Personal Vote-Earning Attributes

This section looks more closely at the running example of personal vote-earning attributes across list types and district magnitudes. Scholars of institutional analysis argue that electoral rules are closely connected to patterns of democratic representation [59]. Some rules favor politicians who provide voters local benefits while others induce politicians to appeal to voters' preferences over national policy goals expressed by political parties [41]. A key conclusion of this literature is that legislators who are in competition with copartisans are likely to seek a personal vote by appealing to factions of party supporters within their districts.

Early studies of the personal vote have generally focused on the linkage between electoral connections and legislative behavior, which states that legislators with personal-vote incentives would engage in constituency-oriented behavior [60,61]. However, relatively few studies examine how an electoral connection affects attributes of legislators such as local origins and locally electoral experience. To understand the connection between the personal vote and legislator attributes, Shugart *et al.* [39] examine how the attributes of legislators vary with electoral rules. Their work asks an important question and focuses on PR systems, in which they argue that the sources of variation in personal vote-earning

attributes lie in information demanded by voters under different list types (closed or open) and different levels of district magnitude. Specifically, in a closed list, voters should rely on the information of candidate attributes when the district magnitude is small. When the magnitude is large, voters would alternatively rely heavily on the cue of party reputation. However, in an open list, voters' demand for information shortcuts is higher, given a specific magnitude, than in a closed list. Moreover, the information demand under open lists increases with magnitude. Based on these theoretical expectations, they hypothesize that the probability that a legislator exhibits PVEA deceases as district magnitude increases in a closed list and increases with district magnitude in an open list.

To test their hypotheses, Shugart *et al.* analyze an original data set based on the biographies of approximately 1,100 legislators in six developed democracies, including three closed-list cases (Norway, Portugal, and Spain) and three open-list cases (Finland, Luxembourg, and Switzerland). They test two personal vote-earning attributes: whether legislators are native to the district they represent, and whether they have prior experience in lower-level elective office within the district. Due to space limitations, only the first one is analyzed here as the outcome variable (`native`), which is coded 1 if the legislator is native to the district in which the legislator is nominated, and 0 otherwise.

The two main explanatory variables measure list type (`open`) and district magnitude (`logM`), respectively. The former is a dummy variable that takes the value of 1 if the list is open and 0 if it is closed; the latter is the decimal logarithm of district magnitudes. What is of most interest is that list type and district magnitude interact to affect the probability that a legislator exhibits PVEA. Therefore, an interaction term between the list type variable and the district magnitude variable is included in the model.

Moreover, Shugart *et al.* argue that district nativity and previous electoral experience in lower office are more important to rookie legislators than they are to veterans. A legislator can engage in personal-vote seeking behavior once they have been elected. For legislators who have won their first race, their attributes related to the district are critical. A rookie legislator is measured here by a dummy variable (`rookie`), which is coded 1 if a legislator is at the first term, and 0 otherwise. It is expected that the effect of magnitude and list type are greater for rookie legislators than for veterans. To test this hypothesis, a second-order interaction model is specified.

6.1. The First-order Interaction Model

This section provides a reanalysis of Shugart *et al.*'s study using the same probit model with interaction terms. Consider the model with a first-order interaction term, in which the interaction between list type and district magnitude is evaluated. The model specification is given as follows:

$$\Pr(\texttt{native} = 1) = \Phi(\beta_0 + \beta_1 \log \texttt{M} + \beta_2 \log \texttt{M} * \texttt{open} + \beta_3 \texttt{open}) \tag{14}$$

The substantive purpose of this model is to test whether the effect of district magnitude on personal vote-earning attributes in the two list types are distinct from each other. One the one hand, it is expected that β_1, the effect of magnitude for closed lists, is negative and statistically significant. On the other hand, $\beta_1 + \beta_2$, the effect of district magnitude for open lists, is expected to be positive and statistically significant. In other words, the interaction effect β_2 is positive and its absolute value is greater than β_1

$(|\beta_2| > |\beta_1|)$. If these two hypotheses are both supported, we should see a difference between β_1 and $\beta_1 + \beta_2$.

Model 1 in Table 1 shows the results of the first-order interaction model. As expected, β_1 is negative and significant at the conventional statistical level (*i.e.*, p-value < 0.05), which means that, in a closed list, the probability that legislators exhibit PVEA decreases with magnitude. Regarding $\beta_1 + \beta_2$, the estimate is positive $(-0.49 + 0.63 = 0.14)$, which is in the expected direction. To see whether the estimate is reliably distinct from zero, the conditional standard error is calculated. [7] The standard error of $\beta_1 + \beta_2$ conditional on open $= 1$ is about 0.27, which fails to give a statistically significant estimate of the effect of district magnitude for open lists. This result suggests that, in an open list, the probability that legislators exhibit PVEA does not increase with magnitude. While the hypothesized effect of magnitude in closed lists is supported, the one in open lists is not, which is inconsistent with the finding in Shugart *et al.* [8]

Table 1. Personal vote-earning attributes: Coefficients estimates.

|	Model 1		Model 2	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Constant (β_0)	0.74	0.14	0.69	0.18
$\log M\ (\beta_1)$	-0.49	0.11	-0.50	0.14
$\log M * \text{open}\ (\beta_2)$	0.63	0.29	0.33	0.39
$\text{open}\ (\beta_3)$	-0.43	0.34	-0.04	0.47
$\text{rookie}\ (\beta_4)$			0.11	0.26
$\log M * \text{rookie}\ (\beta_5)$			0.02	0.20
$\log M * \text{open} * \text{rookie}\ (\beta_6)$			0.36	0.56
$\text{open} * \text{rookie}\ (\beta_7)$			-0.50	0.64
Null Deviance	1493.6 on 1126 d.f.		1473.3 on 1112 d.f.	
Residual Deviance	1471.7 on 1123 d.f.		1448.2 on 1105 d.f.	
AIC	1479.7		1464.2	

Note: Robust standard errors reported.

The left panel of Figure 2 shows the predicted probabilities and 95% confidence intervals of Model 1 in a magnitude range of 5 to 40. Only about 10% of the observations occur in magnitudes out of this range. This also confirms the tabulated results. On one hand, in a closed list, the probability that legislators exhibit PVEA decreases as district magnitude increases. On the other hand, in an open list, the

---

[7] Freely distributed software for the calculation of conditional standard errors is available as an R package, and further technical derivations, replication data, as well as analyses of related specifications are available at the authors' webpage.

[8] In their paper, Shugart *et al.* test for statistical significance of the difference between β_1 and β_2, by means of a χ^2 test, which is statistically significant at $p < 0.05$. Based on this result, they state that the effect of district magnitude matters for both list types and claim that the slopes for district magnitude in the two list types are statistically distinct from each other. However, the χ^2 test does not tell us that $\beta_1 + \beta_2$ is positive and statistically significant: the test only shows that β_1 is distinct from β_2, which is a less important result.

probability that legislators exhibit PVEA does not increase with magnitude, which means that district magnitude does not affect PVEA under open-list systems. [9] Moreover, comparing the left panel of Figure 2 with the right panel of Figure 1, we observe a remarkable difference between the parameter-based interaction and the link function-based interaction. Specifying an interaction term in the model provides a meaningful interpretation of interactive effects in a theoretically-informed manner.

Figure 2. The first-order interaction model. The left panel displays the interactive effect and the right panel displays first differences across districts. That the 95% confidence intervals cover the red line at zero means a lack of statistical significance for first differences.
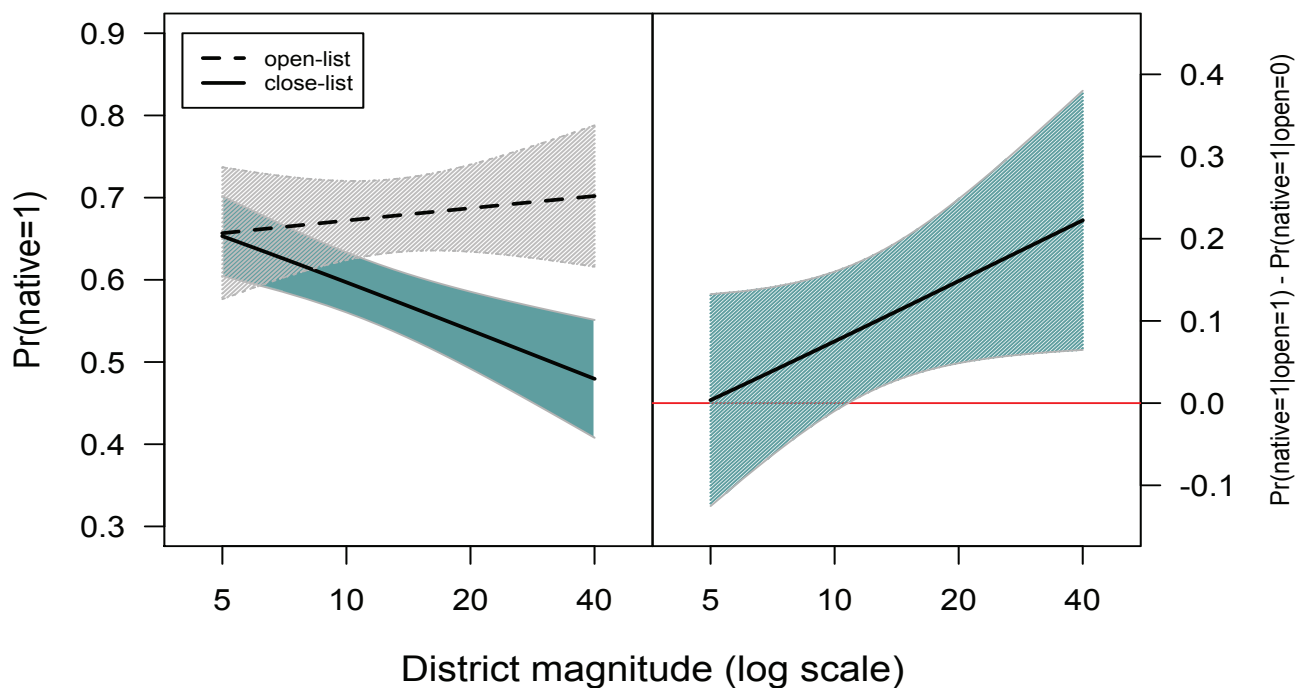

As discussed before, the best way to understand generalized linear model results is by first differences, which provide the effect of the variable of interest conditional on the set value for the other variable that it interacts with. The right panel of Figure 2 shows first differences for PVEA between open and closed lists and their 95% confidence intervals across districts. The values of first differences for PVEA (presented by the black line at the right panel of Figure 2) are derived by subtracting the predicted probability for PVEA under closed lists (presented by the solid line at the left panel of Figure 2) from that under open lists (presented by the dashed line at the left panel of Figure 2). The 95% confidence intervals are calculated accordingly. As can be seen, the first difference probabilities are positive and significant at the conventional statistical level only when $M > 10$. In other words, for district nativity, the effects of list types (closed or open) are indistinguishable at magnitudes smaller than 11.

---

[9] In fact, the lack of statistical significance of the effects for district magnitude can be observed in Figure 2 in Shugart *et al.*'s paper, which is reproduced in the left panel of Figure 2. To see that this is true, notice that the predicted probability for open list systems at a given district magnitude is always covered by the 95% confidence intervals within the given range of district magnitudes, which suggests that, in a magnitude range of 5 to 40, the estimated values of Pr(native) are not different from one another. This cannot be observed by only looking at individual coefficients.

Figure 3 shows first differences and their 95% confidence intervals for the first-order interaction model at three different magnitudes of district ($M = 5, 15$ and 30). First of all, we present the expected differences for the outcome variable across list types, holding the magnitude at the selected levels (*i.e.*, the values along with horizontal arrows). We can see that the probability that legislators exhibit PVEA in open lists is higher than that in closed lists at all three selected magnitudes of district. For instance, given that $M = 15$, a legislator in open lists has a 12% higher probability of exhibiting PVEA than the one in closed lists does. However, the difference between open lists and closed lists is not significant when the magnitude is small. In other words, the attributes of legislators are more important in an open list than in a closed list, but only when the magnitude of district is large enough.

Figure 3. First-order interaction first differences.
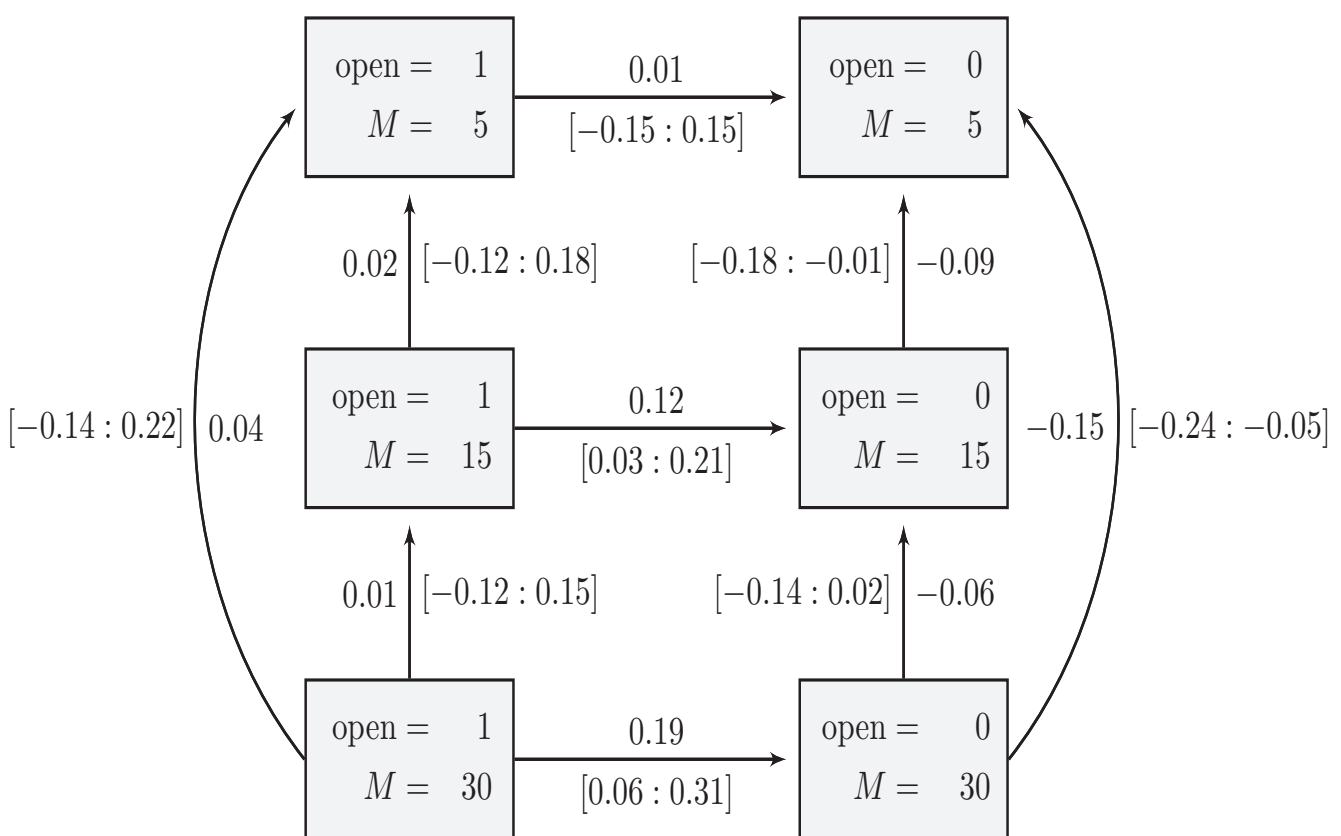
Second, we present comparisons between different levels of district magnitudes given a list type (*i.e.*, the values along with vertical arrows). For example, the first difference probability for increasing the magnitude from 5 to 15 is 0.02 within an interval of -0.12 and 0.18 in open lists and -0.09 within an interval of -0.18 and -0.01 in closed lists. We can see that, under open lists, the values of first differences are positive across magnitudes of district. However, their 95% confidence intervals cover zero, meaning a lack of statistical significance for the effect. Substantively, this means that, under open lists, the probability that a legislator exhibits PVEA does not increase with increasing district magnitude. Regarding closed lists, the values of first differences are negative as we expect. However, the difference is not significant when we increase the magnitude from 15 to 30, which suggests that the probability that a legislator exhibits PVEA decreases as district magnitude increases, but only when the change in

district magnitude is from small to medium. In other words, under closed lists, the decreasing effect of district magnitudes is substantial only in small/medium districts.

6.2. The Second-Order Interaction Model

Now, consider the second-order interaction model, which evaluates the hypotheses that the effect of district magnitude on PVEA is stronger for rookies than it is for veterans under both list types. The model is specified as follows:

$$\Pr(\texttt{native} = 1) = \Phi(\beta_0 + \beta_1 \log \texttt{M} + \beta_2 \log \texttt{M} * \texttt{open} + \beta_3 \texttt{open} + \beta_4 \texttt{rookie} + \beta_5 \log \texttt{M} * \texttt{rookie}$$
$$+ \beta_6 \log \texttt{M} * \texttt{open} * \texttt{rookie} + \beta_7 \texttt{open} * \texttt{rookie}) \tag{15}$$

The coefficient estimates of this model are displayed in Model 2 in Table 1, from which we see that only β_1 is statistically significant at the conventional level (*i.e.*, p-value < 0.05), which is the effect of magnitude for veterans in closed lists. The negative sign indicates that the importance of PVEA for veterans decreases as the district magnitude increases in closed lists.

Figure 4. Second-order interaction first differences for closed lists (`open` $= 0$).
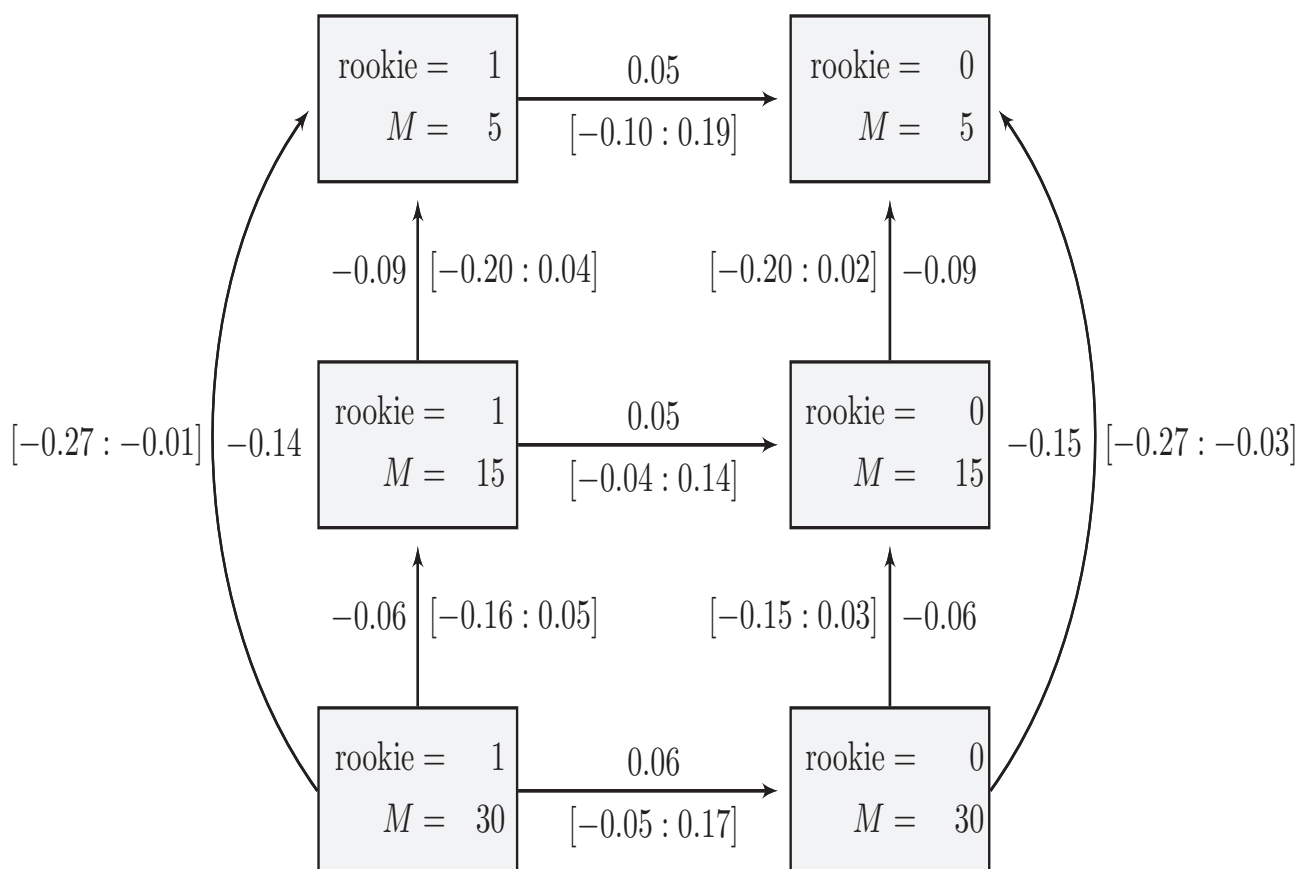
It is expected that the slopes of the effect of district magnitude on PVEA are steeper for rookies than they are for veterans under closed lists ($\beta_1 + \beta_5 < \beta_1$) and under open lists ($\beta_1 + \beta_2 + \beta_5 + \beta_6 > \beta_1 + \beta_2$). To evaluate these expectations, we report first differences for closed lists in Figure 4 and those for open lists in Figure 5. If the inequality $\beta_1 + \beta_5 < \beta_1$ is true, we should observe that the values along with the

horizontal arrows in Figure 4 are negative. As can be seen, these values are positive and not statistically significant at conventional levels, which suggests that the hypothesis is not supported. Moreover, if the inequality $\beta_1 + \beta_2 + \beta_5 + \beta_6 > \beta_1 + \beta_2$ is true, the values along with the horizontal arrows in Figure 5 should be positive. As Figure 5 shows, these expectations are confirmed with one exception when $M = 5$ (their 95% confidence intervals cover zero). Overall, these results suggest that there is no difference between rookies and veterans in exhibiting PVEA under both closed and open lists.

Figure 5. Second-order interaction first differences for open lists (open = 1).
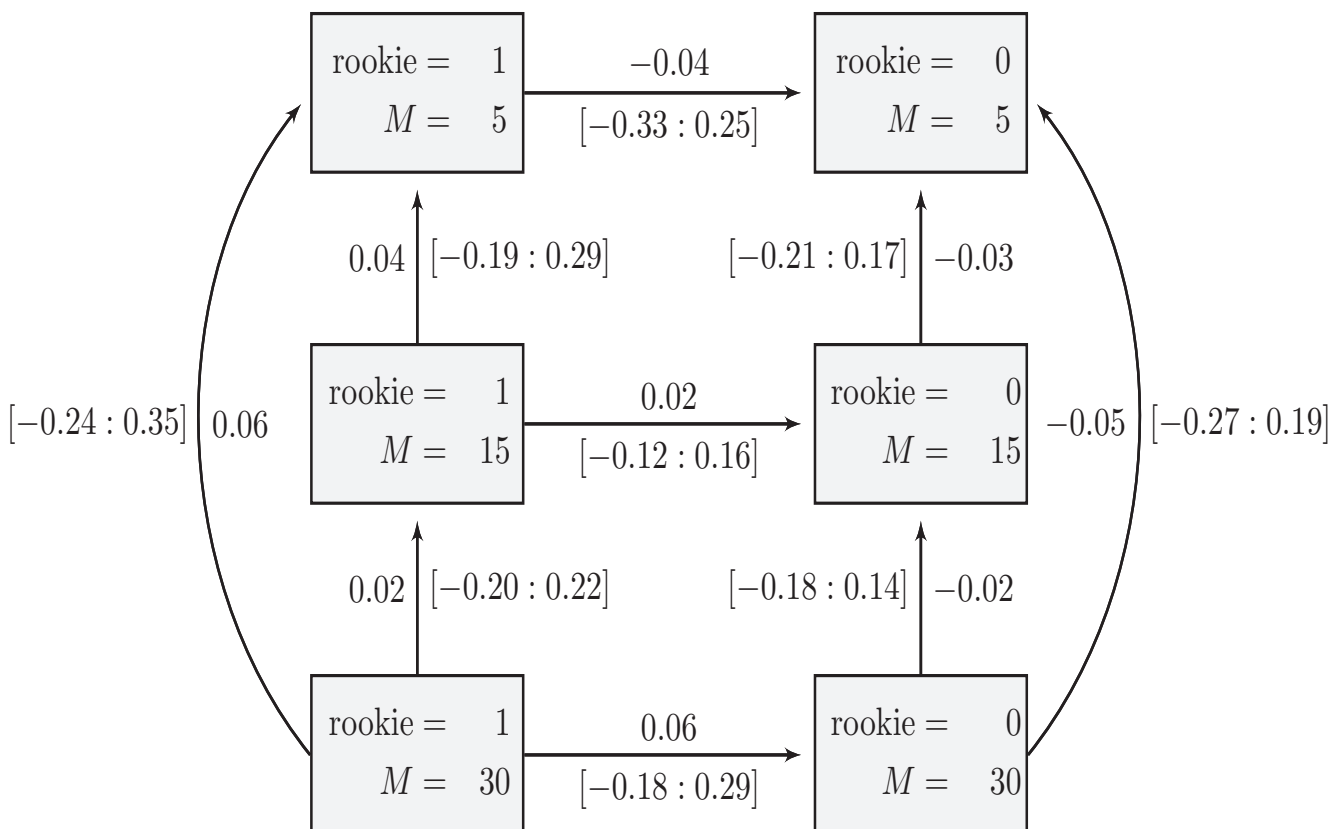
In sum, these findings suggest that personal vote-earning attributes matter more for legislators in open lists than for those in closed lists when district magnitude is large enough. In other words, open lists favor politicians with local connections in high-magnitude districts. Moreover, we do not see a varying effect related to district magnitude in either closed or open lists as Shugart *et al.* claim. Instead, these findings presented here show that district magnitude causes bias against politicians with local connections only in closed lists with small-magnitude districts. Finally, we find no difference between rookies and veterans in exhibiting PVEA in both list types.

7. Conclusions

To a great extent, the mission here has been to build on the excellent, but perhaps forgotten, essay of Friedrich [1]. This article goes beyond Friedrich in that he does not consider the generalized linear model and focuses instead strictly on the interpretation of interactions in the linear model. Friedrich addresses a laundry list of criticisms of the general use of interaction terms, demonstrates why they are overblown,

and then goes on to carefully develop the correct *conditional* nature of interaction coefficients and their standard errors. His work is extended here by building a framework for interpreting and understanding higher-order interaction effects and the more complicated measures of reliability of interaction effects imposed by the generalized linear model. The most important problem corrected by this approach is the impression that the individual standard errors from a regular GLM output capture the full conditional uncertainty of a given interacting model component.

This article argues strongly that interactions are important components of researcher specifications in models of non-normal, non-interval outcome variables, in part because they occur automatically as part of these specifications. When interaction effects are known to exist in a given dataset, they cannot easily be removed by transformation, case-deletion, or a change of method [62], although Tukey's early work [63,64] shows that in the context of analysis of variance such "non-additive" effects can be suitably treated. So rather than considering interactions as annoying features that need to be mitigated, one should view such effects as critical components of the data and specify them in the model in a theoretically-informed manner. Interaction specifications are useful because it is often the case that the full effect of explanatory variables is not purely additive (through the link function). Although there are many ways to specify such models (nonparametrics, polynomials, mixtures, *etc.*), specifying hierarchical interactions is recommended because of the ease of implementation and directly computable solution. Furthermore, specifying interactions as product components is widely applicable across many data forms. Cohen ([65], p. 863) reinforces this point:

> In summary, partial products of variables are correctly interpreted as interactions whatever their level of scaling, whether or not they are correlated, whether or not their means are zero, whether they are observational or experimental in origin, and whether single variables or sets of variables are at issue.

We hope that this discussion improves the specification and interpretation of interactions in the social sciences.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

References

1. Robert J. Friedrich. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26 (1982): 797–833.
2. William D. Berry, Jachueline H. R. Demritt, and Justin Esarey. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54 (2010): 248–66.

3. Frances S. Berry, and William D. Berry. "Specifying a Model of State Policy Innovation." *American Political Science Review* 85 (1991): 573–79.

4. William D. Berry. "Testing for Interaction in Models with Binary Dependent Variables." *Political Methodology Working Paper Archive*, 1999. http://polmeth.wustl.edu/mediaDetail.php?docId=210.

5. Howard Frant. "Specifying A Model of State Policy Innovation." *American Political Science Review* 85 (1991): 571–73.

6. Gary King. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33 (1989): 762–84.

7. Thomas Brambor, William Roberts Clark, and Matt Golder. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (2006): 63–82.

8. Cindy D. Kam, and Robert J. Franzese, Jr. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis.* Ann Arbor: University of Michigan Press, 2007.

9. Jonathan Nagler. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38 (1994): 230–55.

10. Kim Fridkin Kahn, and Patrick J. Kenney. "A Model of Candidate Evaluations in Senate Elections: The Impact of Campaign Intensity." *The Journal of Politics.* 59 (1997): 1173–205.

11. Dean Lacy, and Phillip Paolino. "Downsian Voting and the Separation of Powers." *American Journal of Political Science* 42 (1998): 1180–99.

12. Bruce E. Moon, and William J. Dixon. "Politics, the State, and Basic Human Needs: A Cross-National Study." *American Journal of Political Science* 29 (1985): 661–94.

13. William Roberts Clark, Michael J. Gilligan, and Matt Golder. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14 (2006): 311–31.

14. R. A. Fisher, and W. A. Mackenzie "Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties." *Journal of the Agricultural Society* 13 (1923): 311–20.

15. D. J. Finney. "Main Effects and Interactions." *Journal of the American Statistical Association* 43 (1948): 566–71.

16. William H. Greene. *Econometric Analysis.* Upper Saddle River: Prentice Hall, 2003, Chapter 4.

17. Ronald J. Wonnacott, and Thomas H. Wonnacott. *Econometrics.* New York: John Wiley & Sons, 1970.

18. Paul D. Allison. "Testing for Interaction in Multiple Regression." *American Journal of Sociology* 83 (1977): 144–53.

19. John Fox. *Applied Regression Analysis, Linear Models, and Related Methods.* Thousand Oaks: Sage, 1997, Chapter 14.

20. George E. P. Box, and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* Oxford: John Wiley & Sons, 1987.

21. André I. Khuri, and John A. Cornell. *Response Surfaces: Designs and Analyses.* New York: Marcel Dekker, 1987.

22. John A. Cornell, and Douglas C. Montgomery. "Fitting Models to Data: Interaction Versus Polynomial? Your Choice!" *Communications in Statistics–Theory and Methods* 25 (1996): 2531–55.

23. J. A. Nelder. "A Reformulation of Linear Models" (with discussion). *Journal of the Royal Statistical Society, Series A* 140 (1977): 48–76.

24. D. R. Cox. "Interaction." *International Statistical Review* 52 (1984): 1–31.

25. Phillip Karl Wood, and Paul Games. "Rationale, Detection, and Implications of Interactions Between Independent Variables and Unmeasured Variables in Linear Models." *Multivariate Behavioral Research* 25 (1990): 295–311.

26. James Jaccard, and Choi K. Wan. "Measurement Error in the Analysis of Interaction Effects between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches." *Psychological Bulletin* 117 (1995): 348–57.

27. Irva Hertz-Picciotto, Allan H. Smith, David Holtzman, Michael Lipsett, and George Alexeeff. "Synergism Between Occupational Arsenic Exposure and Smoking in the Induction of Lung Cancer." *Epidemiology* 3 (1992): 23–31.

28. Susan F. Assmann, David W. Hosmer, Stanley Lemeshow, and Kenneth A. Mundt. "Confidence Intervals for Measures of Interaction." *Epidemiology* 7 (1996): 286–90.

29. David W. Hosmer, and Stanley Lemeshow "Confidence Interval Estimation of Interaction." *Epidemiology* 3 (1992): 452–56.

30. Neil Pearce. "Analytical Implications of Epidemiological Concepts of Interaction." *International Journal of Epidemiology* 18 (1989): 976–80.

31. Janet M. Box-Steffensmeier, and Christopher J. W. Zorn. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45 (2001): 972–88.

32. Brad T. Gomez, and J. Matthew Wilson. "Political Sophistication and Economic Voting in the American Electorate: A Theory of Heterogeneous Attribution." *American Journal of Political Science* 45 (2001): 899–914.

33. Diana Evans. "Oil PACs and Aggressive Contribution Strategies." *The Journal of Politics* 50 (1988): 1047–56.

34. Jan E Leighley, and Arnold Vedlitz. "Race, Ethnicity, and Political Participation: Competing Models and Contrasting Explanations." *The Journal of Politics* 61 (1999): 1092–114.

35. Dan Reiter, and Allan C. Stam III. "Democracy, War Initiation, and Victory." *American Political Science Review* 92 (1998): 377–89.

36. Kenneth J. Rothman. *Modern Epidemiology.* Boston: Little Brown & Co., 1986.

37. Ludwig Fahrmeir, and Gerhard Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer, 2001.

38. P. McCullagh, and J. A. Nelder. *Generalized Linear Models.* New York: Chapman & Hall/CRC, 1989.

39. Matthew S. Shugart, Melody E. Valdini, and Kati Suominen. "Looking for Locals: Voter Information Demands and Personal Vote-earning Attributes of Legislators under Proportional Representation." *American Journal of Political Science* 49 (2005): 437–49.

40. Gary W. Cox. *Making Votes Count: Strategic Coordination in the World's Electoral Systems.* New York: Cambridge University Press, 1997.

41. John M. Carey, and Matthew S. Shugart. "Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas." *Electoral Studies* 14 (1995): 417–39.

42. P. J. Green. "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives." *Journal of the Royal Statistical Society, Series B* 46 (1984): 149–92.

43. Guido del Pino. "The Unifying Role of Iterative Generalized Least Squares in Statistical Algorithms." *Statistical Science* 4 (1989): 394–408.

44. Gary King. *Unifying Political Methodology.* New York: Cambridge University Press, 1989.

45. James Jaccard, Robert Turrisi, and Choi K. Wan. *Interaction Effects in Multiple Regression.* Thousand Oaks: Sage, 1990, 41.

46. Neil H. Timm. *Applied Multivariate Analysis.* New York: Springer-Verlag New York, Inc., 2002.

47. David Edwards. "Hierarchical Interaction Models." *Journal of the Royal Statistical Society, Series B* 52 (1990): 3–20.

48. Gary King, Michael Tomz, and Jason Wittenberg. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2000): 341–55.

49. Jan de Leeuw. "Discussion of the Papers by Edwards, and Wermuth and Lauritzen." *Journal of the Royal Statistical Society, Series B* 52 (1990): 21–50.

50. Leo A. Goodman. "Interactions in Multidimensional Contingency Tables." *Annals of Mathematical Statistics* 35 (1964): 632–46.

51. Leona S. Aiken, and Stephen G. West. *Multiple Regression: Testing and Interpreting Interactions.* Thousand Oaks: Sage, 1991.

52. Ralph L. Rosnow, and Robert Rosenthal. "Definition and Interpretation of Interaction Effects." *Psychological Bulletin* 105 (1989): 143–46.

53. Stephen E. Fienberg. *The Analysis of Cross-classified Categorical Data.* Cambridge: MIT Press, 1980.

54. Henry Oliver Lancaster. *The Chi-squared Distribution.* New York: John Wiley & Sons, 1969.

55. Poul Thyregod, and Henrik Spliid. "A Hypothesis of No Interaction in Factorial Experiments with a Binary Response." *Scandinavian Journal of Statistics* 18 (1991): 197–209.

56. Hiroyuki Uesaka. "Test for Interaction Between Treatment and Stratum with Ordinal Responses." *Biometrics* 49 (1993): 123–29.

57. M. E. O'Neill. "The Distribution of Higher-Order Interactions in Contingency Tables." *Journal of the Royal Statistical Society, Series B* 42 (1980): 357–65.

58. Ralph L. Rosnow, and Robert Rosenthal. "If You're Looking at the Cell Means, You're Not Looking at *Only* the Interaction (Unless All Main Effects Are Zero)." *Psychological Bulletin* 110 (1991): 574–76.

59. G. Bingham Powell, Jr. "Political Representation in Comparative Politics." *Annual Review of Political Science* 7 (2004): 273–96.

60. Richard F. Fenno, Jr. *Home Style: House Members in Their Districts.* Boston: Little Brown & Co., 1978.

61. David R. Mayhew. *Congress: The Electoral Connection.* New Haven: Yale University Press, 1974.

62. G. H. Freeman. "The Analysis and Interpretation of Interactions." *Statistician* 12 (1985): 3–10.

63. John W. Tukey. "One Degree of Freedom for Non-Additivity." *Biometrics* 5 (1949): 232–42.

64. Peter G. Moore, and John W. Tukey. "Query 112." *Biometrics* 10 (1954): 562–68.

65. Jacob Cohen. "Partialed Products *Are* Interactions: Partialed Powers *Are* Curve Components." *Psychological Bulletin* 85 (1978): 856–66.