TEACHER'S CORNER

A Tutorial on Testing, Visualizing, and Probing an Interaction Involving a

Multicategorical Variable in Linear Regression Analysis

Andrew F. Hayes and Amanda K. Montoya

The Ohio State University

Department of Psychology

Abstract

Empirical communication scholars and scientists in other fields regularly use regression models to test moderation hypotheses. When the independent variable *X* and moderator *M* are dichotomous or continuous, the practice of testing a linear moderation hypothesis using regression analysis by including the product of *X* and *M* in a model of dependent variable *Y* is widespread. But many research designs include *multicategorical* independent variables or moderators, such as in an experiment with three or more versions of a stimulus where participants are randomly assigned to one of them. Researchers are less likely to receive training about how to properly test a moderation hypothesis using regression analysis in such a situation. In this tutorial, we explain how to test, visualize, and probe interactions involving a multicategorical variable using linear regression analysis. While presenting and discussing the fundamentals—fundamentals that are not software specific—we emphasize the use of the PROCESS macro for SPSS and SAS, as it greatly simplifies the computations and potential for error that exists when doing computations by hand or using spreadsheets based on formulas in existing books on this topic. We also introduce an iterative computational implementation of the Johnson-Neyman technique for finding regions of significance of the effect of a multicategorical independent variable when the moderator is continuous.

A Tutorial on Testing, Visualizing, and Probing an Interaction Involving a

Multicategorical Variable in Linear Regression Analysis

Communication scholars are interested in establishing the effects of communication on a variety of social and behavioral outcomes, such as participation in the political process, happiness and satisfaction in one's relationships with others, or willingness to engage in a particular form of behavior such as getting vaccinated or protecting oneself from cancer or other negative health outcomes, among many other things. They are also interested in the boundary conditions of such effects, meaning when, or for whom, or in what context communication in its many forms has an effect and when it does not, or when its effect is strong or weak, positive or negative. Indeed, it could be argued that the search for moderators of a communication effect is as important when developing and testing theory as is establishing whether or not an effect exists in the first place (c.f., Eveland, 1997).

Variable M is said to *moderate* X's effect on Y if the effect of X on Y depends on M. Moderation is also known as *interaction*. That is, if M moderates X's effect on Y, then X and M interact in their influence on Y (though some make distinctions between moderation and interaction, we do not here). Research that tests a moderation or interaction hypothesis focuses on examining the *contingencies* or boundary conditions of an effect. Establishing such boundary conditions—the conditional nature of a phenomenon—is important in both basic and applied research. For example, establishing that health messages differing in their emotional appeal to change an unhealthy behavior have different effects on perceivers who differ in certain ways (e.g., age, gender, ethnicity, or various psychological or cognitive factors) helps advocates of healthy behavior tailor messages to different audiences while at the same time informing theory about how people process or interpret health-related messages.

Communication scholars, and indeed researchers in most any empirical field, typically are first exposed to the concept of moderation or interaction in the context of coursework on analysis of variance, where interaction manifests itself in the form of *inconsistency in simple effects.* Two simple effects are inconsistent if the effect of one categorical factor on some dependent variable differs across levels of a second categorical factor. Later, researchers fortunate enough to take a good regression course in or acquire one of the many books on data analysis that discuss interactions in regression models (e.g., Cohen, Cohen, West, & Aiken, 2003; Darlington & Hayes, 2017; Hayes, 2005, 2013) learn about testing a moderation hypothesis involving continuous variables, or between a continuous and a dichotomous variable, using linear regression analysis. Most typically, this is accomplished by including the product of X and M in a regression model along with X and M themselves. This allows the effect of X on Y to vary linearly with M.

In our experience, communication scholars and other social scientists are less likely to receive much if any training on how to properly test a moderation hypothesis using regression analysis when one of the variables is multicategorical, meaning categorical with at least three categories. Importantly, many research designs involve multicategorical variables, such as when an experimenter creates or collects three or more versions of a communication-related stimulus that vary on some manipulated dimension or feature (e.g., Feldman & Hart, 2016; Hurley, Jensen, Weaver, & Dixon, 2015; Niederdeppe, Shapiro, Kim, Bartolo, & Porticella, 2014), or when the effects of communication are measured in people who differ in (for example) ethnicity or political party identification (e.g., Habush, Warren, & Benuto, 2012).

To be sure, researchers are exposed to the analysis of multicategorical variables in analysis of variance classes, but they are often left unsure how to test, visualize, or probe a

moderation effect when one of the variables involved in the interaction is continuous. In that situation, the temptation is strong to categorize the continuous variable by artificially slicing the continuum up into groups (using a median split or other form of dichotomization or trichotomization) so that factorial analysis of variance, a procedure many researchers are more familiar with, can be used. Categorization of continuous variables is rarely a good idea (Hayes, 2005; Irwin & McClelland, 2002; Kuss, 2013; Maxwell & Delaney, 1993; Rucker, McShane, & Preacher, 2015; Veiel, 1988), nor is it necessary. Alternatively, a researcher might dichotomize a multicategorical variable by collapsing groups, thereby allowing him or her to use a single crossproduct between X and M in a regression model to estimate interaction. Although this simplifies the analysis, it disrespects the design, throws away distinctions between conditions or groups made at the design phase, and may mask important differences between groups that exist.

In this tutorial, we discuss how to estimate, visualize, and probe interactions involving a multicategorical and a continuous variable in a regression model. Our objective is to familiarize researchers with the basic statistical principles and show how these principles are implemented both in an ordinary regression routine and in the PROCESS macro for SPSS and SAS, a convenient regression-based analytical tool that takes much of the computational burden off the researcher's shoulders. After first discussing how to represent a multicategorical variable in a regression model, we discuss the model comparison approach to testing interaction and show how to visualize a regression model with an interaction involving a multicategorical variable. We then discuss pick-a-point and Johnson-Neyman procedures for probing an interaction, including omnibus approaches and approaches that focus on pairwise comparisons between groups. New to this treatment of the topic is the introduction of an easy-to-use computational tool (OGRS) that finds Johnson-Neyman regions of significance for the effect of a multicategorical variable that

interacts with a continuous variable. We end with various miscellaneous topics such as the inclusion of covariates, assumptions, and multicategorical moderators.

Throughout this tutorial, we use the term *independent variable* or *focal predictor* to refer to the variable X whose effect on an outcome Y is of interest. The moderator variable M, by contrast, is the variable that in some way modifies or influences the size of the independent or focal predictor's effect on Y. This does not mean that M may not be or can't be under the control of an experimenter and thus independently manipulated in some fashion. Ultimately, the symmetry property of interactions (Darlington & Hayes, 2017; Hayes, 2013) tells us that one person's independent variable or focal predictor could be another person's moderator when interpreting the same model, depending on how the question is phrased. Furthermore, sometimes it is helpful to examine a model and probe an interaction twice, flipping the roles of focal predictor and moderator (see e.g., Berry, Golder, & Milton, 2012). But when we say independent variable or focal predictor, we mean the variable whose effect on Y is being estimated. It will always be denoted X in our discussion.

Working Example

We illustrate the principles and procedures relying on data from a study of video game effects conducted by Gabbiadini, Riva, Andrighetto, Volpato, and Bushman (2016) and publicly available through the Open Source Framework archive at https://osf.io/hu85t/. The participants in this study ($n = 155$) were teenagers and young adults (referred to henceforth as "adolescents"), both male and female, varying in age between 15 and 20 years old. Each participant was randomly assigned to play one of three categories of video games that varied in violence and sexist content. The control game contained no violence or sexist content, whereas the other two games contained either violence without sexist content (referred to as the "violence-only"

condition) or both violent content as well as content that included male characters playing

dominant roles over submissive, sexualized female characters (the "violent-sexist" condition).

After 30 minutes of game play, the participants responded to a series of questions used to

quantify their endorsement of masculine beliefs. These included perceptions that it is acceptable

for a man to use any means necessary to convince a woman to have sex, and that boys should be

encouraged to display their physical prowess. This variable—*masculine beliefs*—is the

dependent variable Y in all analyses we report below. Scores on this scale can vary between 1

and 7. In most of this tutorial, our independent variable or focal predictor X is video game

played (i.e., level of violent, sexist content), and the moderator M is age of the participant.

Although age is measured coarsely (years) and thus can take only 6 values in the data, in reality

age is a continuum and is only forced to be discrete through the measurement system used. Thus,

we treat age as a continuum in the analyses we report.

A one-way analysis of variance on masculine beliefs reveals a significant effect of video

game content on such beliefs, $F(2,152) = 3.335, p = .038$. Those who played the violent sexist

game reported higher masculine beliefs on average ($\bar{Y}_3 = 3.025, SD = 0.725$) than did those who

played the violence-only game ($\bar{Y}_2 = 2.778, SD = 0.732$). Those who played the control no

violence game reported the lowest average level of masculine beliefs ($\bar{Y}_1 = 2.628, SD = 0.865$).

But as we will show, the pattern of differences between groups in masculine beliefs following

game play depends on the age of the player. Before discussing how to test and probe an

interaction involving such a multicategorical variable and a continuous moderator, we address

how to represent a multicategorical variable in a regression model.

Representing a Multicategorical Variable in a Regression Model

As discussed in many treatments of regression analysis (e.g., Cohen, Cohen, West, &

Aiken, 2003; Darlington & Hayes, 2017), a multicategorical variable with k categories can

be used as a predictor in a regression model if it is properly represented with $k - 1$

regressors coding the groups represented by the variable. There are many ways this can be

done; only three are described here. Consider a multicategorical variable with $k = 3$ categories,

such as in the video game study. Table 1 shows three ways one might choose to represent three

groups using a group coding system. For all these methods, we construct two variables D_1 and

D_2. The pattern of values for D_1 and D_2 represents the group in which a case resides.

For instance, using *indicator* or *dummy coding* when $k = 3$, $D_1 = D_2 = 0$ for all cases

in the first group, $D_1 = 1$ and $D_2 = 0$ for all cases in the second group, and $D_1 = 0$ and

$D_2 = 1$ for all cases in the third group. The group with zeros on D_1 and D_2 is known

as the *reference group* or *baseline category*. Using this coding system for a three group

categorical variable, in a linear regression model of the form

$$\hat{Y} = b_0 + b_1 D_1 + b_2 D_2 \tag{1}$$

b_1 estimates the mean difference in Y between groups 1 and 2 (i.e., $\bar{Y}_2 - \bar{Y}_1$) and b_2

estimates the mean difference in Y between groups 1 and 3 (i.e., $\bar{Y}_3 - \bar{Y}_1$). If additional

variables are included in Equation 1 as covariates, b_1 and b_2 represent estimated difference

between *adjusted means*, meaning estimated differences in Y between groups holding the

covariate(s) constant. Standard errors, t- and p-values, and confidence intervals available in

regression output provide inference about differences between the means.

To illustrate, Table 2 contains a regression model estimating masculine beliefs from

game condition using an indicator coding system. In the data, the three conditions are held in a

variable named *game* (1 = no violence, 2 = violence-only, 3 = violent-sexist) and *mbelief* is the

participant's score on the measure of masculine beliefs.[1] We used the no-violence condition as the reference ($D_1 = D_2 = 0$), with D_1 coding the violence-only condition ($D_1 = 1$, $D_2 = 0$) and D_2 coding the violent-sexist condition ($D_1 = 0$, $D_2 = 1$). In SPSS, the syntax below generates the indicators codes:

```
compute d1 = (game = 2).
compute d2 = (game = 3).
```

The model accounts for just under 4% of the variance in masculine beliefs, with an F-ratio and p-value that is equivalent to the F-ratio and p-value from the one-way ANOVA, $R^2 = .042$, $F(2,152) = 3.335$, $p = .038$. The regression coefficient for D_1 is $b_1 = 0.150$, which is equal to the mean difference between the violence-only and no violence condition ($\overline{Y}_2 - \overline{Y}_1 = 2.778 - 2.628 = 0.150$). The regression coefficient for D_2 is $b_2 = 0.397$, which is equal to the mean difference between the violent-sexist condition and the no violence condition ($\overline{Y}_3 - \overline{Y}_1 = 3.025 - 2.628 = 0.397$). In a model with no covariates, as this one, the regression constant b_0 is the mean of Y for the reference group, which in this case is the no violence condition, and so $b_0 = 2.628 = \overline{Y}_1$.

Notice that when this model is applied to generate estimates of Y, it generates three estimates that correspond to the groups means. For those assigned to the control no violence condition, $D_1 = D_2 = 0$, and the model generates $\hat{Y} = 2.628 + 0.150(0) + 0.397(0) = 2.628$, which is the mean for the no violence condition. For those in the violence-only condition, $D_1 = 1$ and $D_2 = 0$, and so $\hat{Y} = 2.628 + 0.150(1) + 0.397(0) = 2.778$, which is the mean for the violence-only condition. Finally, for those in the violent-sexist condition ($D_1 = 0$, $D_2 = 1$), $\hat{Y} = 2.628 + 0.150(0) + 0.397(1) = 3.025$, which is the mean for the violent-sexist condition.

Indicator coding is common and perhaps the most frequently used coding system.

---

[1] In the data file as downloaded from the Open Source Framework, video game condition is held in a variable named "condition" and masculine beliefs is named "mas_beli".

Alternative coding systems exist. For example, if the multicategorical variable is an ordinal

variable, sequential coding coding can be useful. Sticking with a $k = 3$ example,

with sequential coding, the ordinally lowest group on the dimension defining the

multicategorical variable is coded with zeros on D_1 and D_2, D_1 is set to one for the other

groups, and D_2 set to one for the ordinally highest group (see Table 1) but zero for the

ordinally middle group. For a complete description of how to generate sequential codes for any

number of groups see Darlington and Hayes (2017). Using this system, regression coefficients

quantify differences on the dependent variable between groups that are ordinally adjacent on the

categorical variable. That is, in a model of the form in Equation 1, b_1 quantifies the estimated

difference in Y between the ordinally middle and the lowest group (i.e., $\overline{Y}_2 - \overline{Y}_1$), and b_2

quantifies the estimated difference in Y between the ordinally highest group and the middle

group (i.e., $\overline{Y}_3 - \overline{Y}_2$).

Although sequential coding was described above as useful when the

multicategorical variable codes an ordinal dimension, it can also be used for a nominal

multicategorical variable if the groups are strategically "ordered" in such a way

during the coding process such that the codes generate comparisons of interest. In the video

game example, if we consider the no violence condition "ordinally lowest," the violence-only

condition as the "middle" condition, and the violent-sexist condition as "ordinally highest," a

sequential coding system will generate regression coefficients that estimate the difference in

masculine beliefs between those the violence-only ($D_1 = 1$, $D_2 = 0$) and no violence conditions

($D_1 = 0$, $D_2 = 0$), and between the violent-sexist ($D_1 = 1$, $D_2 = 1$) and violence-only conditions.

The SPSS code below constructs the sequential codes:

```
compute d1 = (game > 1).
compute d2 = (game > 2).
```

and the model is presented in Table 2. As can be seen, $b_1 = 0.150 = \overline{Y}_2 - \overline{Y}_1 = 2.778 - 2.628$, and

$b_2 = 0.247 = \overline{Y}_3 - \overline{Y}_2 = 3.025 - 2.778$. The regression constant b_0 is $\overline{Y}_1$, the mean of the ordinally

lowest group. This is a mathematically equivalent model, in that it fits just as well as the model

using indicator coding of groups, and it reproduces the group means.

Helmert coding, an alternative coding method for ordinal multicategorical variables,

generates regression coefficients quantifying the difference between means for one group and all

groups ordinally higher on the multicategorical ordinal variable. With three groups, a Helmert

coding scheme is provided in Table 1. For a complete description of how to generate Helmert

codes for any number of groups see Darlington and Hayes (2017). When D_1 and D_2 are regressed

on Y as in Equation 1, b_1 estimates the mean difference in Y between the ordinally lowest group

and the unweighted mean of means for the middle and highest groups [i.e., $0.5(\overline{Y}_2 + \overline{Y}_3) - \overline{Y}_1$], and

b_2 estimates the difference between the means of the middle and ordinally highest groups (i.e.,

$\overline{Y}_3 - \overline{Y}_2$). Readers familiar with contrasts in ANOVA might recognize D_1 and D_2 as representing

two *orthogonal contrasts*.

Table 2 contains the regression model using Helmert coding, which like sequential

coding, can also be used when the group variable is nominal as in the video game example. In

this model, the regression coefficient for D_1 estimates the difference between the unweighted

average masculine beliefs in the violence-only ($D_1 = 1/3$, $D_2 = -1/2$) and violent-sexist ($D_1 = 1/3$,

$D_2 = 1/2$) conditions relative to the no violence condition ($D_1 = -2/3$, $D_2 = 0$), and the regression

coefficient for D_2 estimates the average difference in masculine beliefs between the violence-

only and violent-sexist conditions.[2] The SPSS code below constructs the Helmert codes:

---

[2] When means are combined but unweighted, differences in group sample sizes are ignored in the construction of the combined mean.

```
if (game = 1) d1 = -2/3.
if (game = 1) d2 = 0.
if (game = 2) d1 = 1/3.
if (game = 2) d2 = -1/2.
if (game = 3) d1 = 1/3.
if (game = 3) d2 = 1/2.
```

and as can be seen in Table 2, $b_1 = 0.274$, which is $0.5(\overline{Y}_3 + \overline{Y}_2) - \overline{Y}_1 = 0.5(2.778 + 3.025) -$

2.628. And as discussed above, $b_2 = 0.247$, or $\overline{Y}_3 - \overline{Y}_2 = 3.025 - 2.778$. The regression constant

b_0 is the unweighted mean of the three group means: $(\overline{Y}_1 + \overline{Y}_2 + \overline{Y}_3)/3 = 2.811$. This model fits

the same as the models using indicator or sequential coding and also reproduces the group

means.

We have shown that each of these approaches to coding groups generates equivalent

models as gauged by model fit and the estimates of Y they generate. The test of model fit, in the

form of an F-ratio and p-value for R^2, is the same as the F-ratio and p-value from a one-way

ANOVA comparing the three group means. This test can be construed as test on the fit of the

model, measured with R^2, or, alternatively, as a test of the null hypothesis that both of the

regression coefficients for the D_1 and D_2 variables are equal to zero. That is, under the null

hypothesis of no group differences in means, we would expect both regression coefficients to be

equal to zero regardless of coding scheme. So rejection of the null that the "true" R^2 equals zero

is equivalent to rejecting the null hypothesis that all the regression coefficients for the variables

coding group are equal to zero. This will be important later when we discuss probing an

interaction involving a multicategorical independent variable.

Testing Interaction Involving a Multicategorical Independent Variable

It is well known by communication scholars and other social scientists that interaction

between two continuous variables (or between two dichotomous variables, or between a

dichotomous and a continuous variable) can be tested using regression analysis. If the

independent variable is X and the moderator is M, then in a model of the form

$$\hat{Y} = b_0 + b_1 X + b_2 M + b_3 XM \tag{2}$$

b_3 estimates how much the effect of X on Y changes as M changes by one unit. This can be seen

most clearly by expressing Equation 2 in an equivalent form

$$\hat{Y} = b_0 + (b_1 + b_3 M)X + b_2 M \tag{3}$$

where $b_1 + b_3 M$ is the *conditional effect* of X, which we denote $\theta_{X \rightarrow Y}$ (consistent with the notation

in Hayes, 2013, and Darlington & Hayes, 2017). That is,

$$\hat{Y} = b_0 + \theta_{X \rightarrow Y} X + b_2 M \tag{4}$$

where $\theta_{X \rightarrow Y} = b_1 + b_3 M$. If b_3 is zero, then X's effect on Y is linearly unrelated to M, and Equation

3 reduces to $Y = b_0 + b_1 X + b_2 M$, meaning $\theta_{X \rightarrow Y} = b_1$ and so X's effect on Y is not dependent on

M. Rather, it is a constant.

A test of moderation of the effect of X on Y by M is typically undertaken in one of two

ways. Under the null hypothesis of no linear interaction between X and M, the product of XM

should be ignored when attempting to estimate Y. So a null hypothesis test that the coefficient for

the product term equals zero in Equations 2 or 3 is a test of linear moderation of X's effect on Y

by M. Alternatively, a model comparison approach can be used. This involves assessing the fit of

two models of Y, one with and one without XM in the model. Although the R^2 for the model with

XM will descriptively fit better (as adding a variable to a model almost always increases R^2), a

formal test of significance for the change in R^2 should be conducted. A statistically significant

increase in R^2 when XM is added to a model that contains X and M is affirmative evidence for

moderation of the effect of X on Y by M. These are mathematically equivalent tests and will

generate the same *p*-value for the test of the null hypothesis of no linear interaction between *X* and *M* (see Hayes, 2013, for a discussion).

Before discussing a test of moderation when *X* is multicategorical, it is worth pointing out a frequent confusion. Many researchers mistakenly interpret b_1 in Equation 2 as the "main effect" or "average effect" of *X* on *Y*, aggregating across values of *M*, or as the effect of *X* when "controlling for *M* and the interaction." This misunderstanding has been clarified by many (Hayes, 2013; Hayes, Glynn, & Huge, 2012; Irwin & McClelland, 2001; Spiller, Fitzsimons, Lynch, & McClelland, 2013), however we find it pertinent to repeat it here. In Equation 2, b_1 is a *conditional* effect of *X,* not a main or average effect of *X,* nor is it the effect of *X* controlling for *M* and *XM*. Rather, it estimates the effect of *X* on *Y* when *M* = 0, which is seen most clearly in Equation 3. This estimate and corresponding hypothesis test or confidence interval may or may not have any substantive interpretation, depending on whether zero exists on the measurement scale for *M*. This will be important later when discussing probing an interaction involving a multicategorical *X*.

If *X* is a multicategorical variable coding groups and the goal is to assess the moderation of the *X* effect on *Y* by *M*, the multicategorical variable must first be coded appropriately, such as discussed in the prior section. Once this is accomplished, regress *Y* on the *k* − 1 group codes D_j representing the multicategorical *X* variable, *M*, and the *k* − 1 products involving D_j and *M*. For instance, with *k* = 3 groups, and thus two variables D_1 and D_2 coding the groups,

$$\hat{Y} = b_0 + b_1 D_1 + b_2 D_2 + b_3 M + b_4 D_1 M + b_5 D_2 M \qquad (5)$$

This model allows *X*'s effect on *Y* to depend linearly on *M*. More specifically, it allows the differences between means represented by the coding system used to vary linearly with *M*. Recall from the prior section that when *k* = 3, the regression coefficients for D_1 and D_2 represent

differences between means or groups of means. In Equation 5, these differences are now linear

functions of M, as can be seen by writing Equation 5 in an equivalent form

$$\hat{Y} = b_0 + (b_1 + b_4 M)D_1 + (b_2 + b_5 M)D_2 + b_3 M \tag{6}$$

That is, "the effect of D_1" on Y is $b_1 + b_4 M$, and "the effect of D_2"on Y is $b_2 + b_5 M$. But because

D_1 and D_2 represent X, we can say instead that X's effect on Y is a function of M. Using the

notation introduced earlier, Equation 6 can be expressed as

$$\hat{Y} = b_0 + \theta_{D_1 \to Y}D_1 + \theta_{D_2 \to Y}D_2 + b_3 M \tag{7}$$

where $\theta_{D1 \to Y}$ and $\theta_{D2 \to Y}$ are the conditional effects of D_1 and D_2 on Y (and thus, the conditional

effect of X) defined by the functions $b_1 + b_4 M$ and $b_2 + b_5 M$, respectively.

An omnibus null hypothesis test of no interaction between X and M requires a model

comparison approach. This is accomplished by comparing the fit of this model (in Equation 5) to

a constrained model that fixes b_4 and b_5 to zero. This constrained model looks like

$$\hat{Y} = b_0 + b_1 D_1 + b_2 D_2 + b_3 M \tag{8}$$

which is equivalent to Equations 5 and 6 with the constraint imposed that $b_4 = b_5 = 0$. The fit of

these two models can be compared by computing the difference between their squared

multiple correlations, $\Delta R^2 = R_2^2 - R_1^2$, where R_2^2 and R_1^2 are the squared multiple correlations

for the linear moderation model (Equations 5 or 6) and the constrained model (Equation 8)

respectively.

In its most general form, when X has k categories represented by $k - 1$ codes, under the

standard assumptions of regression and the null hypothesis of no interaction between M and a

multicategorical X, the difference between the fit of the model without the $k - 1$ products (R_1^2 for

the model from Equation 8) and the model with the $k-1$ products (R_2^2 for the model from

Equation 5 or 6) can be converted to an F-ratio

$$F = \frac{df_{residual}(R_2^2 - R_1^2)}{(k-1)(1 - R_2^2)}$$ (9)

that follows the $F(k-1, df_{residual})$ distribution, where $df_{residual}$ is the residual degrees of freedom

from the model expressed in Equation 8. A sufficiently small p-value leads to the inference that

X's effect on Y depends linearly on M. Thankfully, the outcome of this test does not depend on

the coding system used for representing the k groups that constitute X. Whether you use indicator

coding, sequential coding, Helmert coding, or some other legitimate system, the F-ratio and p-

value will be the same. Note that this test does *not* require centering of any of the variables used

to construct the products, contrary to widely-believed myth (see e.g., Hayes, 2013).

This test can be conducted with any statistical program capable of conducting

hierarchical entry of predictors to a regression model or that can test constraints on subsets

of regressors. To illustrate, we examine whether the effect of video game content on

endorsement of masculine beliefs depends on player age. So age is the moderator M, video game

condition is the independent variable X, and the dependent variable remains masculine beliefs.

Because the test of interaction is invariant to method of coding the three conditions, we use

indictor coding of groups with the no violence game as the reference. Later we use Helmert

coding, for reasons discussed then.

The SPSS code below constructs two products between indicator variables held in D_1 and

D_2 (with the control no violence game as the reference) and age and conducts the test of

interaction between game played and age:

```
compute d1age = d1*age.
compute d2age = d2*age.
regression/statistics defaults cha/dep=mbelief/method=enter d1 d2 age
```

```
/method=enter d1age d2age.
```

The resulting model (listed as "Model 1" in SPSS output) that fixes the effect of video game (X)

on masculine beliefs (Y) to be independent of age (M) (i.e., without the two products between the

two D codes and M) is

$$\hat{Y} = 2.055 + 0.146D_1 + 0.451D_2 + 0.033M \tag{10}$$

with $R^2 = 0.044$. For the model with the two products representing the moderation of the effect of

video game play linearly by age (listed as "Model 2" in SPSS output),

$$\hat{Y} = -1.639 + 5.354D_1 + 8.230D_2 + 0.247M - 0.301D_1M - 0.475D_2M \tag{11}$$

and $R^2 = 0.092$. The difference between these (provided in SPSS output) is $\Delta R^2 = 0.048$, $F(2,$

$149) = 3.935$, $p = .022$. Thus, the effect of video game content on masculine beliefs depends on

age of the player.

Again, the outcome of this test is not dependent on how the groups are coded. Had we

used sequential or Helmert coding of conditions rather than indicator coding, we would get the

same result for the test of interaction. But the regression coefficients would be different

depending on the coding system, as they correspond to different comparisons of group means.

Implementation in PROCESS for SPSS and SAS

In the bulk of the rest of this tutorial, we focus our computational instructions on the use

of the PROCESS macro available for SPSS and SAS, as it conducts this test and has many

additional features for probing and visualizing interactions that simplify what are otherwise

rather tedious tasks described later. PROCESS is introduced and documented in Hayes (2013)

and is freely available from *www.processmacro.org* along with an addendum to the

documentation that describes some of its functions specific to testing interactions involving a

multicategorical variable. We show SPSS commands here. The SAS versions of these commands are similar; see the documentation.[3]

In the PROCESS syntax, the focal predictor must be specified as "X" and the moderator specified as "M". Model 1 is a moderation model with M moderating the effect of X on Y. The PROCESS command that conducts this analysis is

```
PROCESS vars=mbelief game age/x=game/m=age/y=mbelief/mcx=3/plot=1/model=1.
```

There is no need to construct the D_1M and D_2M products prior to analysis, as PROCESS understands that these products are needed to estimate the model and will construct them itself.

The resulting output can be found in Appendix 1. The test of interaction is provided in the section of output labeled "R-square increase due to interaction." As can be seen, PROCESS produces the F-ratio and p-value for the test of interaction as well as the change in R^2 resulting from adding the product terms, $\Delta R^2 = .048$, $F(2,149) = 3.935$, $p = .022$. As the R^2 for the unconstrained model is $R^2 = 0.092$ (as shown in the "Model Summary" section of output), we can infer from this output that the fit of the model without the two product terms is $R^2 = 0.092 - 0.048 = 0.044$.

In this command, the "mcx=3" option instructs PROCESS that the variable listed as X is a multicategorical variable and to use Helmert coding of groups. PROCESS does the Helmert coding automatically and displays the codes toward the top of the output. Because the test of interaction is not dependent on the coding system used, other coding systems implemented in PROCESS could be specified instead (e.g., "mcx=1" specifies indicator coding and "mcx=2" specifies sequential coding; see the documentation). Indeed, observe that the test printed by PROCESS is the same as the test generated using the hierarchical variable entry

---

[3] The RLM macro for SPSS and SAS also has features for estimating, probing, and visualizing regression models with an interaction involving a multicategorical variable. See Darlington and Hayes (2017) for a discussion of the RLM macro and guidance on its use.

method when indicator coding was used. Later we clarify why we chose the Helmert coding option rather than indicator coding. The "plot=1" option produces information useful for visualizing the model, the topic of the next section.

Visualizing the Model

Interaction between two variables can be difficult to interpret without a picture of the model. The regression model can be used to produce an estimate of Y at various values of the independent variable and moderator (and covariates; the inclusion of covariates is discussed later) that can then be plotted in order to better understand the nature of the interaction. This requires choosing a set of combinations of the independent variable and moderator and plugging them into the regression equation. When the independent variable or moderator is multicategorical, the choice is determined for you in part. You would want to use all the possible combinations of codes representing the groups. For instance, with a $k = 3$ category focal predictor or moderator, use the three patterns of D_1 and D_2 that code the groups. Along with these, you would choose values of the other variable within the range of the data. For instance, if your independent variable is multicategorical but the moderator is a dichotomous, choose the two values of the moderator as coded in the data. Or if the moderator is continuous, as in our video game example, you could use values corresponding to various percentiles of the distribution, such as the 25th, 50th, and 75th percentiles. An alternative and common choice is the mean, a standard deviation below the mean, and a standard deviation above the mean of the moderator.[4]

---

[4] If using plus and minus one standard deviation from the mean, be careful to check your data to make sure that both of these are within the range of the available data, otherwise the plot will depict the model in a region of space where you have no data or that may be beyond the scale of measurement of the variables.

In our example, video game is a multicategorical variable represented with D_1 and D_2, and age is treated as a continuum. We might choose to use values of the moderator (age) corresponding to the sample mean age (16.826 years), as well as one standard deviation below the mean (15.577 years) and one standard deviation above the mean (18.075 years). When crossed with the three patterns of D_1 and D_2 coding groups, this produces nine combinations of X and M that, when plugged into the regression equation, yields nine estimate of Y for players of different ages who played different games. For example, when using indicator coding, the regression model in Equation 11 generates an estimate of masculine beliefs for an adolescent of average age in the sample (16.826 years) who played the violence-only game ($D_1 = 1$, $D_2 = 0$) equal to

$$\hat{Y} = -1.639 + 5.354(1) + 8.230(0) + 0.247(16.826) - 0.301(1)(16.826) - 0.475(0)(16.826) = 2.811$$

Notice that this is the same estimate that the regression model using Helmert coding generates. From the PROCESS output in Appendix 1, the regression model using Helmert coding of groups is

$$\hat{Y} = 2.889 + 6.792D_1 + 2.876D_2 - 0.012M - 0.388D_1M - 0.174D_2M \tag{12}$$

which when applied to an adolescent of average age who played the violence-only game ($D_1 = 1/3$, $D_2 = -1/2$) generates

$$\hat{Y} = 2.889 + 6.792(1/3) + 2.876(-1/2) - 0.012(16.826) - 0.388(1/3)(16.826) \tag{13}$$
$$- 0.174(-1/2)(16.826) = 2.811$$

Repeating these computations for many combinations of X and M and then plotting the estimates of Y as a function of X and M in the form of a line graph results in Figure 1. The dots on the lines in Figure 1 are nine estimated values of Y using Equations 11 or 12 for combinations of D_1 and D_2 (video game content) and three values of M.

PROCESS has an option that reduces much of the effort required to generate such a plot, and it does the computations to a much higher degree of precision than you would when calculating the estimates of Y by hand. The "plot=1" option generates the table found in the section of output in Appendix 1 titled "Data for visualizing the conditional effect of X on Y". The SPSS version of PROCESS even writes a program to read the data and generate a plot. Cutting and pasting this into an SPSS syntax file and then executing will produce a scatterplot which can be edited to produce something like Figure 1. Indeed, Figure 1 was generated using PROCESS in this way and then editing the resulting plot before imported it into this manuscript.

Most likely, the first thing that comes to your attention looking at Figure 1 is three lines that differ in slope, meaning that the relationship between age and masculine beliefs appears to vary depending on the video game played. Although there is nothing wrong with this interpretation (given the symmetry property of interaction), it reverses the focal predictor and moderator relative to how the research question was phrased. We earlier asked whether the effect of video game content on masculine beliefs varies depending on the age of the player. Framed in this way, our interest is not in the different slopes of the lines in Figure 1, but rather the gap between the lines at given values of age on the horizontal axis. The gap between the lines conditioned at a specific value of age reflects the differences in masculine beliefs resulting from playing games that differ in content. From this perspective, the interaction between age and video game content reflects the fact that the gaps between the lines vary in distance as a function of age. It appears from Figure 1 that games differing in violent and sexist content have the biggest effect on masculine beliefs among younger adolescents compared to the no violence condition. Among older adolescents, the smaller gaps between the lines reflect a smaller effect of

video game content on masculine beliefs. Making sense of how the gaps between the lines vary with age is accomplished by *probing* the interaction, the topic of the next section.

Probing the Interaction

A visual depiction of an interaction helps to make sense of what the data may be telling you about the phenomenon you are studying. But as the visual depiction is generated from model coefficients, each with sampling error attached to them, this means that the visual depiction is sample specific. We observed looking at a picture of this model that it seems that the effect of video game content on masculine beliefs is bigger among younger adolescents than among older ones, but a different sample of 155 adolescents would certainly yield different model coefficients and therefore a different visual depiction of the model. So how much confidence should we have in specific claims such as this when the entire picture has sampling variance attached to it?

Having established interaction and visualizing the model, it is traditional to further the analysis by "probing" the interaction. This exercise involves dissecting the model by estimating the effect of the independent variable on the dependent variable at one or more values of the moderator, followed by an inference about that effect, thereby allowing for more definitive claims about in what region of the moderator the independent variable X seems to be related to Y and in what region it is not.

Here we discuss two methods for probing an interaction, one that is quite popular, and one that is less so but that is becoming more popular as its implementation has been made easier by various computational aides for popular software and tools available online. Both of these methods require distinguishing between an *omnibus test* of the effect of a multicategorical independent variable and a *pairwise* test. An omnibus test answers the vague question "Does X have an effect on Y?" without providing anything specific about the nature of that effect if it

exists. In contrast, a pairwise test determines whether estimates of Y from the model vary between two specific groups or combination of groups, with those comparisons determined by how the multicategorical independent is represented in the regression model (i.e., how the groups are coded). We address omnibus and pairwise tests for each of the two probing methods we outline.

Pick-a-Point Procedure

The *pick-a-point approach*, also called an *analysis of simple slopes* or a *spotlight analysis*, involves choosing a point on the distribution of the moderator M and estimating whether the groups differ from each other on Y, on average, conditioned on that value of M. Any value or values of M can be chosen. When M is continuous, this exercise would typically be repeated for at least two or three values of M.

Omnibus Inference. We saw that in a model of the form

$$\hat{Y} = b_0 + b_1 D_1 + b_2 D_2 + b_3 M + b_4 D_1 M + b_5 D_2 M \tag{14}$$

we can test for interaction between X and M with a test of the null hypothesis that the regression coefficients for the two products are both equal to zero. This is accomplished by quantifying how much R^2 increases when $D_1 M$ and $D_2 M$ are added to the model and converting this change to an F-ratio using Equation 9, from which a p-value can then be derived to test the null hypothesis of no interaction.

Suppose that instead we conduct a test of the null hypothesis that both of the regression coefficients for D_1 and D_2 are zero. In Equation 14, b_1 estimates the effect of D_1 on Y when $M = 0$, and b_2 estimates the effect of D_2 on Y when $M = 0$, because $\theta_{D_1 \to Y} = b_1 + b_4 M$ and $\theta_{D_2 \to Y} = b_2 + b_5 M$, which reduce to b_1 and b_2 when $M = 0$. But D_1 and D_2 together represent the multicategorical X, so testing that b_1 and b_2 are both zero would be like testing whether X has an

effect on Y when $M = 0$. This would be accomplished by calculating R^2 for the model of Y that includes M, D_1M, and D_2M, and then adding D_1 and D_2 and quantifying the increase in R^2. Using Equation 9, we can convert this difference to a F-ratio and then get a p-value. A sufficiently small p-value leads to a rejection of the null that there is no effect of X on Y when $M = 0$.

But what we want is a test of equality of the estimated group means (i.e., Y estimated in each group) when M is some value we choose rather than zero, since zero may not even be a meaningful value on the measurement scale (for example, in the video game example, $M = 0$ corresponds to a newborn—a player with an age of 0). For instance, we could test whether video game content affects masculine beliefs among adolescents who are "relatively young." We can operationalize "relatively young" however we want. A common procedure is to use a specific value in the distribution of the data such as a standard deviation below the mean, or the 25th percentile. Of course, any other reasonable and justifiable value could be used.

This test is accomplished by first *centering* M around the value chosen by constructing a new variable $M' = M - m$, where m is the value of M chosen for the conditioning. If we define "relatively young" adolescents as those one standard deviation below the mean ($m = 15.577$ or about 15 and half years old), then we construct a new variable $M' = M - 15.577$. Importantly, note that when $M = 15.577$, $M' = 0$. We then estimate two models:

$$\text{Model 1: } \hat{Y} = b_0 + b_3 M' + b_4 D_1 M' + b_5 D_2 M' \tag{15}$$

$$\text{Model 2: } \hat{Y} = b_0 + b_1 D_1 + b_2 D_2 + b_3 M' + b_4 D_1 M' + b_5 D_2 M' \tag{16}$$

and test whether Model 2 fits better than Model 1 using Equation 9 to get the F-ratio for the test. This is a test of the null hypothesis that the regression coefficients for D_1 and D_2 are both zero, which is equivalent to the null hypothesis that the three group means are equal when $M' = 0$. But since $M' = 0$ when $M = m$, we are actually testing the null hypothesis of equality of the estimated

group means when $M = m$, which in our example is 15.577 years old and exactly what we want to test. The outcome of this test will not depend on the system used for coding the groups, nor does it matter that you may not have any cases in the data with a value of M exactly equal to m.

When this test is implemented using the video game data (using Helmert coding of groups as described earlier) Model 1 is

$$\hat{Y} = 2.858 - 0.094M' - 0.090D_1M' - 0.156D_2M' \tag{17}$$

with $R^2 = 0.013$, and Model 2 is

$$\hat{Y} = 2.708 + 0.753D_1 + 0.163D_2 - 0.012M' - 0.388D_1M' - 0.174D_2M' \tag{18}$$

with $R^2 = 0.092$. Model 2 is a better fitting model, $F(2,149) = 6.424$, $p < .022$. So we conclude that among a group of relatively young adolescents (about 15 and a half years old), video game content has an effect on masculine beliefs. Using the equation for Model 2, setting M' to 0 and applying the model for the three combinations of D_1 and D_2 Helmert codes representing the three groups, we can derive that the estimated group means among such a group of adolescents are 2.206, 2.878, and 3.041 for the no violence, violence-only, and violent-sexist conditions, respectively. These are the points on the three lines in Figure 1 when $M = 15.577$.

This procedure can then be repeated for some other values of M. For instance, we could repeat this process for adolescents "relatively moderate" in age, operationalized as the median age in the sample or perhaps the sample mean, and then for adolescents "relatively older", such as a standard deviation above the mean age or perhaps the 75th percentile. When we did so using the mean age ($M = 16.826$ or just under 17 years old) and a standard deviation above the mean age ($M = 18.076$ or about 18 years old), the F-ratios and p-values for the change in R^2 were $F(2,149) = 1.710$, $p = 0.184$, and $F(2,149) = 0.377$, $p = 0.687$, respectively. The estimated means being compared are the points on the line in Figure 1 when $M = 16.826$ and $M = 18.076$. From

these tests, we conclude that among adolescents relatively moderate or older in age, video game content appears not to affect masculine beliefs, on average. The results of this study suggest that video game content affects masculine beliefs only among the relatively younger adolescent players.

PROCESS automatically implements this method when the independent variable is specified as a multicategorical variable using model 1. As can be seen in Appendix 1, PROCESS provides a test of the difference in estimated Y for the three video game conditions at values of the moderator corresponding to a standard deviation below the mean, the mean, and a standard deviation above the mean. PROCESS prints the F-ratio, change in R^2, a p-value, and also the estimates of Y from the model when the moderator is centered around the specific value displayed in the corresponding section of output. The F-ratios and p-values are the same as the F-ratios and p-values we generated in our discussion above. PROCESS also constructs the estimated means conditioned on these values of the moderators, using the procedure described earlier.

If you prefer to use different values of the moderator for conditioning the omnibus test of the effect of the independent variable than PROCESS uses by default, you can do so with a couple of options. You can tell PROCESS to condition on various percentiles of the distribution, or you can choose a specific value of the moderator. See the documentation.

Pairwise Inference. This omnibus inference for the effect of the independent variable at values of M results in only the vague conclusion that there is or is a not a difference between the k groups on Y when M is a specific value. It does not specify which groups differ from which. Given this, one might forgo the omnibus approach entirely and look for differences between specific groups on Y at one or more values of M. Alternatively, one might use the omnibus

inference about the effect of the independent variable at a value of M as a gatekeeper for deciding whether or not to further explore whether specific groups differ from which at that value of M.

In this example, we have used Helmert coding of groups. This was a strategic decision and guided by what seemed like the most meaningful comparisons when examining the effects of content on masculine beliefs. In this study, there are three groups of adolescent participants who played games differing in content. We could have used an indicator coding system with the no violence condition as the reference, with D_1 and D_2 used as indicators for those assigned to the violence-only and violent-sexist conditions, respectively. As discussed earlier, effects for D_1 would then quantify differences in mean masculine beliefs between the no violence and violence-only conditions, and effects for D_2 would quantify differences in mean masculine beliefs between the no violence and violent-sexist conditions. But a more meaningful set of comparisons would be focused on the following questions: Does any type of violent video game content (whether violence-only or violent-sexist) result in an elevation of masculine beliefs relative to no such content? And does adding sexist content to violence produce a difference in masculine beliefs relative to when the content is just violent?

Helmert coding accomplishes this. Using the Helmert coding system described earlier, the effect of D_1 quantifies the difference between the average masculine beliefs of those in the no violence condition and the unweighted average masculine beliefs for those who received *some* violent content (violence-only, or violent-sexist). And the effect of D_2 quantifies the difference between the average masculine beliefs of those who received only violent content and those who received both violent and sexist content. For the omnibus test, the coding makes no difference in the results obtained, but for pairwise tests, the system of coding used determines the information

you can extract from the regression model about relative differences between specific groups or combinations of groups.

The presence of the products involving M and the D variables complicates interpretation of D_1 and D_2's effects. From the PROCESS output in Appendix 1 (also see our discussion resulting in Equation 12) the regression model when using Helmert coding of groups is

$$\hat{Y} = 2.889 + 6.792D_1 + 2.876D_2 - 0.012M - 0.388D_1M - 0.174D_2M \qquad (19)$$

In this model, the effect of D_1 is $\theta_{D1 \to Y} = 6.792 - 0.388M$, and the effect of D_2 is $\theta_{D2 \to Y} = 2.876 - 0.174M$. Both of these differences are a function of \age (M), with b_4 and b_5 quantifying the extent to which age is related to these differences. But do we have evidence that these effects significantly vary with age? Hypothesis tests for the regression coefficients for the two products answer this question. As can be seen in the PROCESS output, b_4 (the coefficient for the D_1M product) is statistically significant, $b_4 = -0.388$, $p = .006$, meaning that age and content, as represented by the D_1 comparison, interact. In other words, the effect of receiving any violence relative to no violence depends on age. However, b_5 (the coefficient for the D_2M product) is not statistically significant, $b_5 = -0.174$, $p = .347$, meaning that age and content, as represented by the D_2 comparison, do not interact. So the effect of receiving sexist content in addition to violence relative to only violent content does not significantly vary with age.

Having established that age is related to the size of the effect of any violent content compared to no violence, we can then probe this interaction by estimating this effect at specific values of age. Earlier we operationalized "relatively young," "relatively moderate" and "relatively older" adolescent players using the mean, a standard deviation below the mean, and a standard deviation above the mean. Sticking with these operationalizations, let's estimate the effect of some violence relative to none among relatively young players. Recall that in Equation

14, b_1 estimates the effect of D_1 on Y when $M = 0$ (because $\theta_{D1 \rightarrow Y} = b_1 + b_4M$, which is b_1 when $M = 0$), but that in Equation 16, b_1 estimates the effect of D_1 on Y when $M' = 0$, where M' is M centered around some value m. So we make $M' = 0$ when $M = 15.577$ by centering M around 15.577 and then estimating Equation 16. Doing so yields

$$\hat{Y} = 2.708 + 0.753D_1 + 0.163D_2 - 0.012M' - 0.388D_1M' - 0.174D_2M' \qquad (20)$$

(recall we already estimated this in the omnibus inference section earlier; see Equation 18). In this model, b_1 estimates the effect of D_1 (i.e., some violent content relative to none) among relatively young players, and we get a standard error, and t- and p-values for this comparison. In this analysis, $b_1 = 0.753$, $p = .002$. So we can say that among relatively young adolescents, some violent video game content results in significantly higher masculine beliefs than does no violent content. Repeating this process, centering M around the mean age and again a standard deviation above the mean age yields $b_1 = 0.269$, $p = .113$ when $M = 17.826$; and $b_1 = -0.215$, $p = 0.388$, when $M = 18.076$. So we see no statistically significant differences in masculine beliefs when there is some violent content compared to none among moderate and older adolescents.

Earlier, we didn't find evidence that age interacts with the effect of violent content alone compared to violence plus sexist content. Typically, without evidence of an interaction, there would be no point to estimating the effect of X (or some subset of X, such as comparisons represented by D_1 or D_2) at different values of M. An investigator who does so might be tempted to interpret a difference in p-values for the conditional effect at different values of M as evidence that these conditional effects differ from each other, which implies interaction. But difference in significance does not imply significantly different (c.f., Gelman & Stern, 2006). It is the test of interaction that provides a formal test of whether conditional effects of X differ across values of moderator M. However, if there is specific point of interest on the moderator continuum, it may

still be informative to estimate the conditional effect of X at that value of M. For instance, we might check whether, for relatively young adolescents (among whom we saw an effect of some violent content), whether adding sexist to violent content results in a difference in masculine beliefs. In the model with M centered around 15.577 (Equation 20), b_2 answers this question. In this model, $b_2 = 0.163$ and is not statistically significant ($p = .478$), meaning no significant difference in masculine beliefs on average among the younger adolescents who played a game with violence only relative to those who played a game with violent and sexist content.

PROCESS automates this whole procedure. Examining the second half of the output in Appendix 1, observe that PROCESS generates estimates of the effect of X at values of M corresponding to a standard deviation below the mean, the mean, and a standard deviation above the mean. The omnibus test of group differences was discussed already. Just above these three sections of output, PROCESS prints b_1 and b_2 after centering M in the fashion just described, along with standard errors, t- and p-values, and confidence intervals. Also provided are the model derived estimates of Y for the three groups conditioned on these values of M. These correspond to the nine estimated values in Figure 1. Options exist for choosing values of M, or using percentiles of the distribution of M for centering. See the documentation.

The Johnson-Neyman Technique

Although the pick-a-point procedure is common and popular, it suffers from a major flaw. Typically, there is little rationale for choosing some values of M rather than others. The choice is usually arbitrary, and yet the choice will affect the resulting tests and inference. Although using the mean, a standard deviation below the mean, and a standard deviation above the mean is a common convention, this is still arbitrary, and no more or less arbitrary than choosing some other values, such as various percentiles of the distribution of the moderator. On

occasion, theory or applied concerns might suggest certain meaningful values of M to choose, but this is rather rare in our experience.

The problem of how to select values of the moderator without relying on arbitrary conventions or haphazard choices is solved through the use of the *Johnson-Neyman* (JN) *technique* (Johnson & Neyman, 1936; Bauer & Curran, 2005) . The JN technique for probing an interaction is like the pick-a-point approach in reverse. Rather than choosing a value or values of M in advance and then estimating and testing X's effect on Y at that value or values, the JN technique derives the value or values along the continuum of M at which point the effect of X on Y is just statistically significant at a specified level of significance, such as .05. The value or values found (if any) demarcate the *region of significance* of the effect of X on Y. With these values derived, one is then in a position to discuss for whom or under what circumstances, as operationalized by M, X exerts an effect on Y and for whom or under what circumstances it does not.

Omnibus Inference. Though historically not widely used, recent implementations of the JN technique on the web (Preacher, Curran, & Bauer, 2006) and in macros for SPSS and SAS (Hayes, 2013; Hayes & Matthes, 2009) when X is dichotomous or a continuum are probably responsible for its recent increase its popularity, including in the field of communication (for recent uses, see e.g., Beullens, Rhodes, & Eggermont, 2016; Nan, Madden, Richards, Holt, Wang, & Tracy, 2016). However, when X is a multicategorical variable, the derivation of regions of significance for an omnibus group comparison is quite a bit more complicated than when X is dichotomous or a continuum. Quite simply, the mathematics is very tedious. When X represents three groups, it involves finding the roots of fourth order polynomial. With four groups, the

polynomial is of the eighth order. With four or more groups, no solution can be derived

analytically.

To deal with this problem, Montoya (2016) produced a computational tool for SPSS and

SAS named OGRS (for Omnibus Groups Regions of Significance) that iteratively searches the

moderator space in search of points along the moderator continuum where the effect of a

multicategorical X on Y transitions between statistically significant and not significant. Using the

bisection method (Burden & Faires, 1985), it breaks the moderator continuum into segments and

calculates the F-ratio and corresponding p-value for a test of the null hypothesis of no group

differences in Y at many values of M within each segment. It searches iteratively within each

segment until if finds an F-ratio for the effect of X on Y at a value of M that is just statistically

significant at a specified α-level or it determines that no such value of M exists in that segment.

It repeats this across the entire observed range of the moderator and reports the values of M

found. So OGRS does not derive the regions of significance analytically, but rather, through an

iterative computational hunt. But tests of its effectiveness described by Montoya (2016) show

that it is usually quick to find these points of transition, and the points it finds when they *can* be

derived analytically correspond very closely to what OGRS reports.

OGRS is publicly available and downloadable from *www.akmontoya.com* along with the

documentation. In our example, the goal is to find the region or regions of significance along the

age continuum for the effect of video game content on masculine beliefs. The OGRS command

in SPSS is

```
ogrs vars=mbelief game age/y=mbelief/x=game/m=age.
```

The resulting output can be found in Appendix 2. As can be seen, OGRS automatically codes the

variable specified as X using indicator coding (as this is an omnibus approach, the solution is not

dependent on the system of coding used), reports the regression model and test of interaction,

prints any values of M found where the F-ratio for the effect of X is just statistically significant at

.05 (the α-level can be changed; see the documentation), and produces a table of many values of

M, the F-ratios for the effect of X at those value of M, corresponding changes in R^2 described

earlier, and p-values. As can be seen, OGRS identifies $M = 16.398$ (a bit under 16 and a half

years old) as a transition point defining a region of significance. The table shows that among

adolescents this old or younger, there is a statistically significant effect of video game content on

masculine beliefs at the $\alpha = .05$ level. But among adolescents older than this, there is no

statistically significant difference in average masculine beliefs resulting from differences in the

content of the video game.

Notice that this result corresponds with our previous findings from the pick-a-point

approach, where we found that among relatively young adolescents ($M = 15.577$) there was an

effect of video game content on masculine beliefs. However, for those individuals with age

greater than 16.398, there were no effects of video game content on masculine beliefs ($M =$

16.826 and $M = 18.076$). But the Johnson-Neyman technique generates results that do not

require an choice of values of M prior to conducting the test. Rather, based on the transition point

identified, we know at what values of M the effect of X is statistically significant and at what

values it is not.[5]

Pairwise Inference. Remember that with k groups, the $k - 1$ D variables coding groups

represent comparisons between estimates of group means, depending on the system used for

coding groups. These differences are functions of the moderator when the $k - 1$ crossproducts

involving the Ds and M are in the model. For instance, with three groups, as discussed earlier,

---

[5] It is important to acknowledge that any point of transition identified by the Johnson-Neyman method is merely a point estimate. As it is derived from regression coefficients and standard errors that are subject to sampling variance, so too is a region of significance identified by this method. In principle, one could construct a confidence interval for a Johnson-Neyman solution, but we are not aware of any published work on this topic.

the conditional effect of D_1 is $\theta_{D_1 \to Y} = b_1 + b_4 M$ and the conditional effect of D_2 is $\theta_{D_2 \to Y} = b_1 +$

$b_5 M$. Both of these depend on M. Different values of M will generate different estimates of mean

differences corresponding to what the D variable codes, and each of these estimates has an

associated p-value for testing the null hypothesis that the conditional effect of that D at that value

of M is different from zero, as discussed earlier.

The JN method can be used for each of the $k - 1$ D variables, or instead only those that

significantly interact with M, to find the value or values of M in the function for the effect of that

D that demarcate regions of significance for the difference coded by the D variable. This can be

done in PROCESS using a simple trick. With an X with k groups and thus $k - 1$ D variables

coding groups, specify one of the D variables as X, and the other $k - 2$ D codes *along with their*

products with M as covariates. When using this approach and the JN method is requested,

PROCESS will find the value of values of M at which $\theta_{D_j \to Y}$ is just statistically different from

zero at the $\alpha = .05$ level (the α-level can be changed; see the documentation). This can be

repeated for different D variables, swapping out which D plays the role of X and which D with

corresponding products with M are used as covariates.

To illustrate, recall that our use of Helmert codes produces two D variables whose

regression coefficients quantify how the masculine beliefs of those in the no violence condition

differ from the unweighted average beliefs of those in the violent game conditions (D_1), and a

second quantifying how much the beliefs of those in the violence only differ from those in the

violent-sexist condition (D_2). Since our earlier analysis revealed no statistically significant

moderation of the difference between violence-only and violent-sexist conditions by age (i.e., no

$D_2 M$ interaction), we will not search for regions of significance for this comparison. But we did

find an interaction between age and the effect of some violence relative the nonviolent game, so

we will apply the JN method to probe this component of the omnibus interaction.

To implement this approach, D_1 and D_2 must be constructed and therefore exist in the

data prior to executing a PROCESS command. Unlike in our earlier example, PROCESS won't

generate these codes for you when using this trick because the mcx option in PROCESS cannot

be used in conjunction with the JN command. So the commands in SPSS that implement this

pairwise Johnson-Neyman approach using Helmert coding of groups are

```
if (game = 1) d1 = -2/3.
if (game = 1) d2 = 0.
if (game = 2) d1 = 1/3.
if (game = 2) d2 = -1/2.
if (game = 3) d1 = 1/3.
if (game = 3) d2 = 1/2.
compute d2age=d2*age.
PROCESS vars=d1 d2 age d2age mbelief/y=mbelief/x=d1/m=age/model=1/jn=1.
```

Observe that we had to first construct and then include the D_2M product in the PROCESS

variable list, but we did not have to do this for D_1M, as PROCESS will generate this product for

itself automatically for Model 1. If we included D_1M in the variable list, a long string of errors

would result because the set of predictors would be singular (i.e., the tolerances of some of the

predictors would be zero).

The PROCESS output generated by this command can be found in Appendix 3. The only

difference between this output and the output in Appendix 1 is the inclusion of the JN table at the

end. As can be seen, PROCESS identifies two ages, 16.657 and 19.880, where the estimated

difference between those in the no violence condition and those in either of the violent game

conditions transitions between significant and not significant at the .05 level. The first region of

significance is ages below just about 16 and a half years old, where the violent games seemed to

result in more masculine beliefs relative to the nonviolent game (because $\theta_{D1 \to Y}$ is positive in this

region). The other region is about age 20 and higher, where the reverse seems to be the case.

However, PROCESS tells us that only 1.29% of the cases in the data are above the age value 19.8804. There are just two 20 year olds in the study, and they are the oldest two participants study. So it would not be prudent to interpret this region given how little data there are in this region of the moderator. The first transition point has about 50% of the data on either side of it, so any inference about the conditional effect of D_1 at values above or below it is more trustworthy.

<div align="center">**Miscellaneous Matters in Conclusion**</div>

In this tutorial we have described the theory and mechanics of testing, visualizing, and probing an interaction between a multicategorical and continuous variable in linear regression analysis. Although all these procedures can be conducted with any OLS regression program, the PROCESS and OGRS macros make the analysis easy and packages the relevant information in one handy output. In this closing section we discuss the inclusion of covariates, when the moderator is dichotomous rather than continuous, multicategorical moderators, and assumptions.

Covariates

To simplify our treatment, we have discussed the estimation and probing of an interaction involving a multicategorical variable using a regression model that includes only the independent variable and the moderator. In practice, tests of moderation are frequently conducted by including additional variables in the model acting as statistical controls or *covariates*. But including covariates does not alter the principles and application we have described here. They can simply be included in the regression model, and almost everything we have discussed still applies. Covariates can be included in a PROCESS and OGRS command by including them in the list of variables in the model. See the documentation.

There is one exception to this. When visualizing an interaction, the estimates of Y using the procedure we described above are generated from the complete regression model. Those computations cannot ignore the covariates. When the model includes covariates, a sensible tradition is to set the covariates to their means when generating the estimates of Y for visualizing the model. The means for the covariates are weighted according to their regression coefficients, and this generates estimates of Y for cases that vary on the independent variable and moderator but that are average on all the covariates. Although it might seem counterintuitive, it is mathematically appropriate to use the mean of a dichotomous covariate in these computations, even when the coding of the dichotomous variable is arbitrary.

A Dichotomous Moderator

The mathematics and procedure we have described for testing and visualizing interaction involving a multicategorical independent variable in this paper apply without modification when the moderator is dichotomous. But the Johnson-Neyman procedure could not be used for probing the interaction because there would be only two discrete values of the moderator. In that case, the pick-a-point procedure involves estimating the effect of X on Y at the two values of the moderator. If the moderator listed in a PROCESS command is dichotomous, PROCESS will automatically produce tests of the interaction and conditional effects of X on Y at each value of the moderator. No further input from the user is required.

When the Moderator is Multicategorical

The symmetry property of interactions means that if X's effect on Y is moderated by M, then M's effect on Y is moderated by X. In our example, the focal predictor X was a multicategorical variable and moderator M was a continuous variable, but the model and procedure for testing interaction would be the same if M were multicategorical and X were

continuous or dichotomous. In that case, M rather than X would be represented by $k - 1$ variables coding group membership. For instance, if M were the three groups coding video game condition, represented with two D variables, and X were age, the model allowing age's relationship with masculine beliefs to vary between the three video game conditions would be

$$\hat{Y} = b_0 + b_1 X + b_2 D_1 + b_3 D_2 + b_4 D_1 X + b_5 D_2 X \qquad (21)$$

We already estimated this model twice above, once using indicator and once using Helmert coding of experimental condition that we were calling X then but are now calling M. Equation 21 can be rewritten in a form that expresses X's effect on Y as a function of the categorical M:

$$\hat{Y} = b_0 + (b_1 + b_4 D_1 + b_5 D_2)X + b_2 D_1 + b_3 D_2$$

meaning that the conditional effect of X is $\theta_{X \to Y} = b_1 + b_4 D_1 + b_5 D_2$ and so it is a function of the categorical M. With estimates of b_1, b_4, and b_5, we can plug the three combinations of D_1 and D_2 representing the three groups into this function to get the slopes of the three lines in Figure 1. For example, when using indicator coding of video game condition with the no violence condition as the reference (as earlier), we have

$$\hat{Y} = -1.639 + 0.247X + 5.354 D_1 + 8.230 D_2 - 0.301 D_1 X - 0.475 D_2 X \qquad (22)$$

(which is Equation 11, but we are swapping the roles of X and M so we exchange symbols in Equation 22) and so the slope of the line in Figure 1 relating age to masculine beliefs among those in the violence only condition ($D_1 = 1$, $D_2 = 0$) is

$$\theta_{X \to Y} = 0.247 - 0.301(1) - 0.475(0) = -0.054 \qquad (23)$$

Comparable computations yield slopes of 0.247 for the no violence condition ($D_1 = D_2 = 0$) and −0.228 for the violent-sexist condition ($D_2 = 1$, $D_2 = 0$). Had we used the regression coefficients from the model with Helmert coding of groups (Equation 12, swapping X and M), we'd get the same conditional effects of age on masculine beliefs in the three conditions. And if you wonder

whether these slopes differ significantly from each other, they do. That test was already

conducted when we tested whether the effect of video game condition on masculine beliefs

differs by age. It does, as the symmetry property tells us that, therefore, the effect of age on

masculine beliefs differs between the three video game conditions.

PROCESS takes care of these computations for you while providing the test of

interaction and also inferential tests for each of these conditional slopes. The command is

`PROCESS vars=mbelief game age/x=game/m=age/y=mbelief/mcm=1/model=1.`

where `mcm=1` tells PROCESS that the moderator M is a multicategorical variable and to use

indicator coding of groups (`mcm=2` generates sequential coding of groups, and `mcm=3` generates

Helmert codes). The output (see Appendix 4) shows that the relationship between age and

masculine beliefs is statistically significant only in the no violence condition.

The regression coefficients for the products and corresponding inferential tests provide

further information. In this model, b_4 and b_5 estimate differences between pairs of slopes. With

the indicator coding system used here, b_4 is the difference between the slopes relating age to

masculine beliefs in the no violence and violence-only condition ($b_4 = -0.301 = -0.054 - 0.247$),

and b_5 is the difference between the slopes in the no violence and violent-sexist conditions ($b_5 =$

$-0.475 = -0.228 - 0.247$). As can be seen in Appendix 4, each of these is statistically significant,

meaning that these pairs of slopes are significantly different from each other. We don't get a test

comparing the slopes relating age to masculine beliefs in the violence-only and violent-sexist

conditions. This test can be conducted by doing the analysis again, changing the reference

category or using a different coding system that produces such a comparison.

Multiple Testing

When conducting an analysis of variance comparing several groups, one frequently

requires multiple hypothesis tests or confidence intervals to test for differences between groups

and find the source of those differences, if any. This raises the specter of Type I error inflation.

This problem, if it is even construed as such, can be exacerbated when moderation is added to

the analysis. In addition to a test of interaction, probing an interaction involving a

multicategorical variable typically requires several additional hypothesis tests. An investigator

who is inclined to worry about reporting false positives might ponder compensating for multiple

tests, such as a Bonferroni correction or using some other approach, prior to accepting the

results. But our position is that replication rather than statistical adjustments to p-values or

confidence intervals is the best way to ensure that findings that result from a many hypothesis

tests can be trusted as reflecting something real about the process or phenomenon being studied.

Linearity, Homoscedasticity, and Other Assumptions

As a regression-based procedure, all the assumptions of ordinary least squares regression

apply; that the effects estimated are linear and the errors in estimation of Y are conditionally

normal, equal in variance (homoscedastic), independent, and centered at zero. The use of

crossproducts to represent interaction also assumes that the moderation is linear, meaning that

the effect of the independent variable on the dependent variable is linearly related to the

moderator. Nonlinear moderation is mathematically possible and is accomplished by changing

the function linking the moderator to the size of the independent variable's effect on Y. See, for

example, Cohen et al. (2003).

For a detailed discussion of regression assumptions, see Berry (1993) and Hayes and

Darlington (2017). Here we address the normality and homoscedasticity assumptions. The

assumption of normality is the least important of these assumptions, as regression is quite robust

to nonnormal errors in estimation.[6] The effects of heteroscedasticty (i.e., violation of homoscedasticity, or equal conditional variance in errors) on tests of interaction have been widely studied (for a review, see Aguinis & Pierce, 1998; more recently, see Overton, 2001, and Sheih, 2009), and it is known that inequality in the variance of the errors in estimation can have adverse effects on tests of interaction, either by lowering power or, more likely, increasing the likelihood of a false positive (i.e., claiming an interaction exists that really does not). Although there are many approaches to modifying estimation or hypothesis tests to deal with this (e.g., weighted least squares), perhaps the simplest is the use of a heteroscedasticity-consistent covariance matrix of the regression parameter estimates. The estimation of the standard errors of regression coefficients as well as the F-ratio in Equation 9 rely on the variance-covariance matrix of the regression coefficients. Various estimators of this matrix exist that can be substituted in the computations (see Hayes and Cai, 2007; Long & Ervin, 2000) that don't rely on homoscedasticity. The net result is a test of interaction that appears to be much less, if at all, affected by heteroscedasticity (see Hayes & Agler, 2014, for a small scale simulation). PROCESS implements one of these—the HC3 estimator—so with PROCESS it is possible to get reasonably valid inference about interaction without assuming equality of the variance of errors in estimation. See the documentation for guidance.

---

[6] Contrary to widespread belief, the normality assumption is about the shape in the errors of estimation of Y, not Y itself or the distribution of the predictor variables.

References

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of

moderating effects of categorical variables: A conceptual review. *Organizational*

Research Methods, *1*, 296–314.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression:

Inferential and graphical techniques. *Multivariate Behavioral Research, 40*, 373–400.

Berry, W. D. (1993). *Understanding regression assumptions*. Thousand Oaks, CA: Sage

Publications.

Berry, W. D., Golder, M., & Milton, D. (2012). Improving tests of theories positing interaction.

Journal of Politics, 74, 653-671.

Beullens, K., Rhodes, N., & Eggermont, S. (2016). Behavioral activation and inhibition as

moderators of the relationship between music video-viewing and joyriding attitudes.

Media Psychology, 19, 181–202.

Burden, R. L., & Faires, J. D. (1985). *Numerical analysis* (3rd ed.). Boston, MA: Prindle, Weber,

& Schmidt.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression and*

correlation for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum

Associates.

Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models: Concepts,*

applications, and implementation. New York, NY: The Guilford Press.

Eveland, W. P. (1997). Interactions and nonlinearity in mass communication: Connecting theory

and methodology. *Journalism and Mass Communication Quarterly, 74*, 400–416.

Gabbiadini, A., Riva, P., Andrighetto, L., Volpato, C., & Bushman, B. J. (2016). Acting like a tough guy: Violent-sexist video games, identification with game characters, masculine beliefs, and empathy for female violence victims. *Plos One, 11*(4), 1–14.

Feldman, L., & Hart, P. S. (2016). Using political efficacy messages to increase climate activism: The mediating role of emotions. *Science Communication, 38*, 99–127.

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *American Statistician, 60*, 328-331.

Habush, A., Warren, C. S., & Benuto, L. (2012). Beauty, ethnicity, and age: Does internalization of mainstream media ideals influence attitudes toward older adults? Sex Roles, 66, 668–676.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

Hayes, A. F. (2005). *Statistical methods for communication science*. New York: Routledge.

Hayes, A. F., & Agler, R. A. (2014). On the standard error of the difference between two independent regression coefficients in moderation analysis: A commentary on Robinson, Tomek, and Schumacker (2013). *Multiple Linear Regression Viewpoints*, *40*(2), 16–27.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*, 709–722.

Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions regarding the interpretation of regression coefficients and hypothesis tests in linear models with interactions. *Communication Methods and Measures, 6*, 1–11.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression. *Behavior Research Methods, 41*, 924–936.

Hurley, R. J., Jensen, J., Weaver, A., & Dixon, T. (2015). Viewer ethnicity matters: Black crime in TV news and its impact on decisions regarding public policy. *Journal of Social Issues, 71*, 155–170.

Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research, 38*, 100–109.

Johnson, P. O., & Neyman, J. (1936). Test of certain linear hypotheses and the application to some statistical problems. *Statistical Research Memoirs, 1*, 57–93.

Kuss, O. (2013). The danger of dichotomizing continuous variables: A visualization. *Teaching of Statistics*, *35*, 78–79.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity-consistent standard errors in the linear regression model. *American Statistician, 54*, 217–224.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*, 181–190.

Montoya, A. K. (2016). *Extending the Johnson-Neyman technique to categorical independent variables: Mathematical derivations and computational tools* (Master's thesis). The Ohio State University.

Nan, X., Madden, K., Richards, A., Holt, C., Wang, M. Q., & Tracy, K. (2016). Message framing, perceived susceptibility, and intentions to vaccinate children against HPV among African American Parents. *Health Communication, 31*, 798–805.

Niederdeppe, J., Shapiro, M. A., Kim, H. K., Bartolo, D., & Porticella, N. (2014). Narrative

persuasion, causality, complex integration, and support for obesity policy. *Health Communications, 29* , 431–444.

Overton, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods*, **6**, 218-233.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*, *25*, 666–678.

Shieh, G. (2009). Detection of interactions between a dichotomous moderator and a continuous predictor in moderated multiple regression with heterogeneous error variance. *Behavior Research Methods, 41*, 61–74.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research, 50*, 277–288.

Veiel, H. O. F. (1988). Base-rates, cut-points and interaction effects: The problem with dichotomized continuous variables. *Psychological Medicine, 18*, 703–710.

Appendix 1

The PROCESS output below is from a model with masculine beliefs as the dependent variable *Y*, game content as a multicategorical independent variable *X*, and age as moderator *M*. Helmert coding is used for coding the three came content conditions.

PROCESS vars=mbelief game age/x=game/m=age/y=mbelief/mcx=3/plot=1/model=1


```
*************** PROCESS Procedure for SPSS Release 2.16 *****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
     Documentation available in Hayes (2013). www.guilford.com/p/hayes3

**********************************************************************
Model = 1
     Y = mbelief
     X = game
     M = age

Sample size
        155

Coding of categorical X variable for analysis:
   game     D1      D2
   1.00   -.67     .00
   2.00    .33    -.50
   3.00    .33     .50

**********************************************************************
Outcome: mbelief

Model Summary
           R          R-sq         MSE           F          df1         df2           p
       .3028         .0917        .5789       3.0072      5.0000    149.0000        .0129

Model
               coeff         se           t           p         LLCI        ULCI
constant      2.8892       1.1572      2.4967       .0136        .6025      5.1759
age           -.0116        .0704      -.1651       .8691       -.1508       .1275
D1            6.7918       2.3228      2.9240       .0040       2.2020     11.3816
D2            2.8762       2.9793       .9654       .3359      -3.0110      8.7633
int_1         -.3876        .1382     -2.8048       .0057       -.6607      -.1145
int_2         -.1742        .1845      -.9438       .3468       -.5388       .1905

Product terms key:
  int_1    :        D1         X         age
  int_2    :        D2         X         age

R-square increase due to interaction:
    R2-chng           F          df1         df2           p
      .0480       3.9346       2.0000    149.0000        .0216

**********************************************************************
```

```
Conditional effect of X on Y at values of the moderator:

Moderator value:
age     15.5767

          Coeff          se           t           p          LLCI         ULCI
D1       .7536         .2353       3.2033       .0017        .2887       1.2185
D2       .1634         .2296        .7114       .4780       -.2904        .6171

Test of equality of conditional means at this value of the moderator
     R2-chng           F           df1          df2           p
       .0783        6.4251       2.0000     149.0000        .0021

Estimated conditional means at this value of the moderator
        game          yhat
       1.0000        2.2057
       2.0000        2.8777
       3.0000        3.0410
-------------
Moderator value:
age     16.8258

          Coeff          se           t           p          LLCI         ULCI
D1       .2694         .1691       1.5931       .1133       -.0647        .6035
D2      -.0542         .2465       -.2198       .8263       -.5413        .4330

Test of equality of conditional means at this value of the moderator
     R2-chng           F           df1          df2           p
       .0208        1.7101       2.0000     149.0000        .1844

Estimated conditional means at this value of the moderator
        game          yhat
       1.0000        2.5140
       2.0000        2.8105
       3.0000        2.7563
-------------
Moderator value:
age     18.0750

          Coeff          se           t           p          LLCI         ULCI
D1      -.2148         .2479       -.8667       .3875       -.7047        .2750
D2      -.2717         .4184       -.6494       .5171      -1.0985        .5551

Test of equality of conditional means at this value of the moderator
     R2-chng           F           df1          df2           p
       .0046         .3770       2.0000     149.0000        .6866

Estimated conditional means at this value of the moderator
        game          yhat
       1.0000        2.8223
       2.0000        2.7433
       3.0000        2.4716
```

Moderator values are the sample mean and plus/minus one SD from mean

****

Data for visualizing conditional effect of X on Y
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/game age mbelief.
BEGIN DATA.

```
     1.0000     15.5767     2.2057
     2.0000     15.5767     2.8777
     3.0000     15.5767     3.0410
     1.0000     16.8258     2.5140
     2.0000     16.8258     2.8105
     3.0000     16.8258     2.7563
     1.0000     18.0750     2.8223
     2.0000     18.0750     2.7433
     3.0000     18.0750     2.4716
```

END DATA.
GRAPH/SCATTERPLOT=age WITH mbelief BY game.

******************** ANALYSIS NOTES AND WARNINGS ************************

Level of confidence for all confidence intervals in output:
     95.00

Appendix 2

Output from the OGRS macro for finding the regions of significance of the effect of game content on masculine beliefs.

ogrs vars=mbelief game age/y=mbelief/x=game/m=age.

```
******************* OGRS Procedure for SPSS Version 1.1 *******************

                        Written by Amanda Montoya

*************************************************************************

Variables:
X = game
M = age
Y = mbelief

Dummy Variable Coding Scheme:
 game   D1   D2
    1    1    0
    2    0    1
    3    0    0

Sample size:
  155

*************************************************************************
Outcome: mbelief

Model Summary
         R        R-sq         F         df1         df2           p
      .3028       .0917     3.0072      5.0000    149.0000       .0129

Model
            coeff         SE          t          p        LLCI        ULCI
constant   6.5912     2.4858     2.6515      .0089      1.6792     11.5033
D1        -8.2299     3.0587    -2.6907      .0079    -14.2739     -2.1859
D2        -2.8762     2.9793     -.9654      .3359     -8.7633      3.0110
age        -.2279      .1587    -1.4359      .1531      -.5416       .0857
Int1        .4747      .1892     2.5096      .0132       .1009       .8485
Int2        .1742      .1845      .9438      .3468      -.1905       .5388

Interactions:
 Int1    =        D1       X        age
 Int2    =        D2       X        age

R-square increase due to interaction(s):
    R2-chng         F         df1         df2           p
      .0480     3.9346      2.0000    149.0000       .0216

*********************** JOHNSON-NEYMAN TECHNIQUE ************************

Moderator value(s) defining Johnson-Neyman boundaries of significance:
    16.3983
```

```
Conditional effect of X on Y at values of the moderator (M)
      age     R2-chng          F           p
    15.0000     .0841       6.8995       .0014
    15.2500     .0840       6.8934       .0014
    15.5000     .0804       6.5975       .0018
    15.7500     .0717       5.8800       .0035
    16.0000     .0585       4.8004       .0095
    16.2500     .0446       3.6553       .0282
    16.3983     .0373       3.0568       .0500
    16.5000     .0328       2.6927       .0710
    16.7500     .0234       1.9214       .1500
    17.0000     .0152       1.2481       .2900
    17.2500     .0081        .6634       .5166
    17.5000     .0034        .2756       .7595
    17.7500     .0021        .1713       .8427
    18.0000     .0037        .3067       .7363
    18.2500     .0069        .5686       .5676
    18.5000     .0106        .8653       .4230
    18.7500     .0140       1.1495       .3196
    19.0000     .0171       1.4037       .2489
    19.2500     .0198       1.6248       .2004
    19.5000     .0221       1.8154       .1663
    19.7500     .0241       1.9798       .1417
    20.0000     .0259       2.1220       .1234


***************************************************************************
```

Appendix 3

The code below implements a trick using PROCESS to generate regions of significance for a pairwise group pairwise when the independent variable is multicategorical.

```
if (game = 1) d1 = -2/3.
if (game = 1) d2 = 0.
if (game = 2) d1 = 1/3.
if (game = 2) d2 = -1/2.
if (game = 3) d1 = 1/3.
if (game = 3) d2 = 1/2.
compute d2age=d2*age.
PROCESS vars=d1 d2 age d2age mbelief/y=mbelief/x=d1/m=age/model=1/jn=1.
```

```
************** PROCESS Procedure for SPSS Release 2.16 *****************

        Written by Andrew F. Hayes, Ph.D.      www.afhayes.com
    Documentation available in Hayes (2013). www.guilford.com/p/hayes3

***********************************************************************
Model = 1
    Y = mbelief
    X = d1
    M = age

Statistical Controls:
CONTROL= d2        d2age

Sample size
       155

***********************************************************************
Outcome: mbelief

Model Summary
          R         R-sq        MSE          F         df1         df2           p
      .3028        .0917       .5789      3.0072      5.0000    149.0000       .0129

Model
            coeff         se          t          p        LLCI        ULCI
constant    2.8892     1.1572     2.4967      .0136       .6025      5.1759
age         -.0116      .0704     -.1651      .8691      -.1508       .1275
d1          6.7918     2.3228     2.9240      .0040      2.2020     11.3816
int_1       -.3876      .1382    -2.8048      .0057      -.6607      -.1145
d2          2.8762     2.9793      .9654      .3359     -3.0110      8.7633
d2age       -.1742      .1845     -.9438      .3468      -.5388       .1905

Product terms key:

 int_1    d1          X        age

R-square increase due to interaction(s):
        R2-chng          F        df1        df2          p
int_1      .0480      7.8669     1.0000    149.0000       .0057
```

```
**********************************************************************


Conditional effect of X on Y at values of the moderator(s):
      age      Effect        se         t          p        LLCI       ULCI
   15.5767      .7536      .2353     3.2033      .0017      .2887     1.2185
   16.8258      .2694      .1691     1.5931      .1133     -.0647      .6035
   18.0750     -.2148      .2479     -.8667      .3875     -.7047      .2750


Values for quantitative moderators are the mean and plus/minus one SD from mean.
Values for dichotomous moderators are the two values of the moderator.

******************** JOHNSON-NEYMAN TECHNIQUE *************************


Moderator value(s) defining Johnson-Neyman significance region(s)
      Value     % below    % above
    16.6568    45.8065    54.1935
    19.8804    98.7097     1.2903


Conditional effect of X on Y at values of the moderator (M)
      age      Effect        se         t          p        LLCI       ULCI
   15.0000      .9772      .2963     3.2976      .0012      .3916     1.5627
   15.2500      .8802      .2687     3.2765      .0013      .3494     1.4111
   15.5000      .7833      .2428     3.2268      .0015      .3036     1.2630
   15.7500      .6864      .2193     3.1307      .0021      .2532     1.1197
   16.0000      .5895      .1990     2.9623      .0036      .1963      .9827
   16.2500      .4926      .1831     2.6905      .0079      .1308      .8544
   16.5000      .3957      .1727     2.2912      .0234      .0544      .7369
   16.6568      .3349      .1695     1.9760      .0500      .0000      .6698
   16.7500      .2988      .1689     1.7692      .0789     -.0349      .6325
   17.0000      .2019      .1720     1.1733      .2425     -.1381      .5418
   17.2500      .1050      .1818      .5771      .5647     -.2544      .4643
   17.5000      .0080      .1973      .0407      .9676     -.3818      .3979
   17.7500     -.0889      .2172     -.4092      .6830     -.5180      .3403
   18.0000     -.1858      .2404     -.7728      .4409     -.6608      .2893
   18.2500     -.2827      .2661    -1.0623      .2898     -.8085      .2432
   18.5000     -.3796      .2936    -1.2928      .1981     -.9598      .2006
   18.7500     -.4765      .3225    -1.4775      .1417    -1.1138      .1608
   19.0000     -.5734      .3524    -1.6271      .1058    -1.2698      .1230
   19.2500     -.6703      .3831    -1.7498      .0822    -1.4274      .0867
   19.5000     -.7673      .4144    -1.8515      .0661    -1.5861      .0516
   19.7500     -.8642      .4462    -1.9368      .0547    -1.7458      .0175
   19.8804     -.9147      .4629    -1.9760      .0500    -1.8294      .0000
   20.0000     -.9611      .4783    -2.0092      .0463    -1.9063     -.0159


******************** ANALYSIS NOTES AND WARNINGS *************************


Level of confidence for all confidence intervals in output:
    95.00
```

Appendix 4

The PROCESS output below is from a model with masculine beliefs as the dependent variable *Y*, game content as a multicategorical moderator variable *M*, and age as the independent variable *X*. Indicator coding is used for coding the three game content conditions.

PROCESS vars=mbelief game age/x=game/x=age/y=mbelief/mcm=1/model=1.

```
************** PROCESS Procedure for SPSS Release 2.16 ******************

          Written by Andrew F. Hayes, Ph.D.        www.afhayes.com
     Documentation available in Hayes (2013). www.guilford.com/p/hayes3

*************************************************************************
Model = 1
    Y = mbelief
    X = age
    M = game

Sample size
       155

Coding of categorical M variable for analysis:
   game     D1     D2
   1.00    .00    .00
   2.00   1.00    .00
   3.00    .00   1.00

*************************************************************************
Outcome: mbelief

Model Summary
          R         R-sq        MSE           F        df1        df2            p
      .3028        .0917      .5789      3.0072     5.0000   149.0000        .0129

Model
               coeff          se           t           p        LLCI        ULCI
constant     -1.6387      1.7822      -.9195       .3593     -5.1603      1.8829
age            .2468       .1029      2.3984       .0177       .0435       .4501
D1            5.3537      2.4235      2.2091       .0287       .5650     10.1425
D2            8.2299      3.0587      2.6907       .0079      2.1859     14.2739
int_1         -.3006       .1394     -2.1555       .0327      -.5761      -.0250
int_2         -.4747       .1892     -2.5096       .0132      -.8485      -.1009

Product terms key:
 int_1     :        D1          X         age
 int_2     :        D2          X         age

R-square increase due to interaction:
    R2-chng          F         df1         df2            p
     .0480      3.9346      2.0000    149.0000        .0216

*************************************************************************
```

```
Conditional Effect of Focal Predictor in Groups Defined by the Moderator
Variable:
      game      coeff        se         t         p      LLCI      ULCI
    1.0000      .2468     .1029    2.3984     .0177     .0435     .4501
    2.0000     -.0538     .0941    -.5713     .5686    -.2397     .1322
    3.0000     -.2279     .1587   -1.4359     .1531    -.5416     .0857

******************** ANALYSIS NOTES AND WARNINGS ************************

Level of confidence for all confidence intervals in output:
    95.00
```

Table 1

Indicator, Sequential, and Helmert Coding of Three Groups

Group	Indictor		Sequential		Helmert	
	D_1	D_2	D_1	D_2	D_1	D_2
1 (Control)	0	0	0	0	–2/3	0
2 (Violence-only)	1	0	1	0	1/3	–1/2
3 (Violent-sexist)	0	1	1	1	1/3	1/2

Table 2

Regression Models of the Form $\hat{Y} = b_0 + b_1D_1 + b_2D_2$ When Estimating Masculine Beliefs as a

Function of Video Game Content Using the Three Systems for Coding Groups in Table 1

		Coeff.	se	t	p
Indicator Coding					
	b_0	2.628	0.107	24.497	<.001
	b_1	0.150	0.150	1.005	.316
	b_2	0.397	0.155	2.564	.011
Sequential Coding					
	b_0	2.628	0.107	24.497	<.001
	b_1	0.150	0.150	1.005	.316
	b_2	0.247	0.153	1.614	.109
Helmert Coding					
	b_0	2.811	0.062	45.161	<.001
	b_1	0.274	0.132	2.078	.039
	b_2	0.247	0.153	1.614	.109

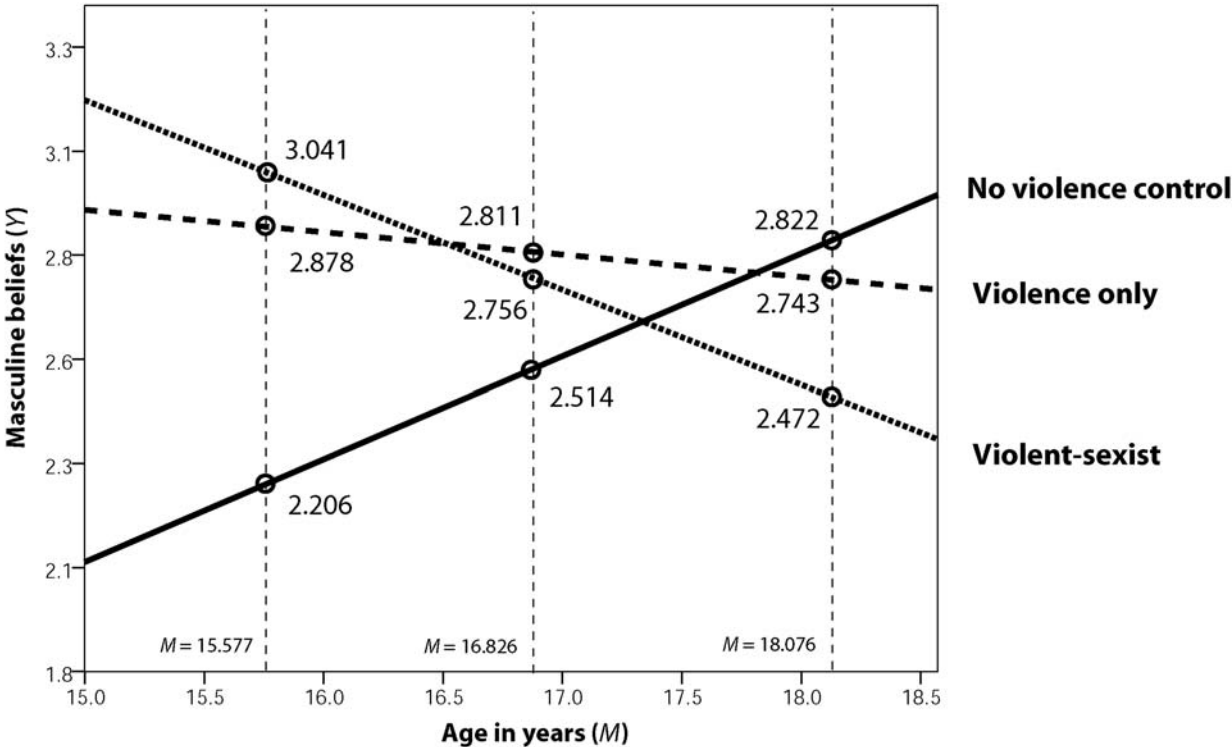For all three models, $R^2 = .042$, $F(2,152) = 3.335$, $p = .038$.

Figure 1. A visual depiction of the interaction between video game content and player age.