Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression

STEPHEN A. SPILLER
GAVAN J. FITZSIMONS
JOHN G. LYNCH, JR.
GARY H. MCCLELLAND

November 15, 2012

Stephen A. Spiller is Assistant Professor of Marketing, Anderson School of Management, UCLA, email: stephen.spiller@anderson.ucla.edu. Gavan J. Fitzsimons is R. David Thomas Professor of Marketing and Psychology, Fuqua School of Business, Duke University, email: gavan@duke.edu. John G. Lynch, Jr. is the Ted Anderson Professor at the Leeds School of Business, University of Colorado-Boulder and the Director of the Center for Research on Consumers' Financial Decision Making, email: john.g.lynch@colorado.edu. Gary H. McClelland is Professor of Psychology and Faculty Fellow at the Institute of Cognitive Science, University of Colorado-Boulder, email: Gary.McClelland@colorado.edu. Fitzsimons, Lynch, and McClelland are listed alphabetically and contributed equally. The authors thank Rick Staelin, Carl Mela, and Wagner Kamakura for asking questions that motivated this article and Ajay Abraham, Philip Fernbach, Yoosun Hann, Ji Hoon Jhang, Christina Kan, Peggy Liu, Matthew Philp, Adriana Samper, Julie Schiro, Scott Wallace, Elizabeth Webb, Hillary Wiener, and the review team for constructive comments, as well as seminar participants at the University of Colorado. Any errors are the authors'.

Spotlights, Floodlights, and the Magic Number Zero:

Simple Effects Tests in Moderated Regression

Abstract

It is common for authors discovering a significant interaction of a measured variable X with a manipulated variable Z to examine simple effects of Z at different levels of X. These "spotlight" tests are often misunderstood even in the simplest cases, and it appears that consumer researchers are unsure how to extend them to more complex designs. We explain the general principles of spotlight tests, show that they rely on familiar regression techniques, and provide a tutorial showing how to apply these tests across an array of experimental designs. Rather than following the common practice of reporting spotlight tests at one standard deviation above and below the mean of X, we recommend that when X has focal values, researchers report spotlight tests at those focal values. When X does not have focal values, we recommend researchers report ranges of significance using a version of Johnson and Neyman's (1936) test we call a "floodlight".

Keywords: Moderated regression; Spotlight analysis; simple effects tests

Most marketing and consumer behavior papers reporting experiments test for interactions between two or more variables. Authors may follow up an interaction of two variables with "simple effects" tests of the effect of one variable at different levels of another – called "conditional effects" by econometricians. They may follow up an interaction of three variables with tests of "simple interactions" of two variables at a level of a third variable or "simple-simple" effects of one variable at chosen levels of the other two (cf. Keppel and Wickens 2004).

This paper presents a tutorial on the analysis of simple effects tests in designs when one or more of the interacting variables are continuous and quantitative rather than categorical. Researchers primarily trained in using ANOVA frameworks for experimental designs often struggle when following up interactions where a continuous variable interacts with one or more categorical variables, where the appropriate analysis takes place in the framework of a moderated regression. In a review of volume 48 of the *Journal of Marketing Research* (*JMR*) and volume 38 of the *Journal of Consumer Research* (*JCR*), we found that the reported moderated regression analyses were often "correct" but not optimally performed, and in many cases were simply incorrect. We observe similar small and large errors in other social sciences. In this paper, we identify the most common misunderstandings and we provide a simple framework for conducting these analyses.

Consider a fictional extension of McFerran, Dahl, Fitzsimons, and Morales's (2010) study of the effect of social influence on consumption. The authors proposed that individuals' own consumption behavior is anchored on the quantity taken by others in their environment, but they adjust their consumption based on whether others around them belong to an aspirational or dissociative group. That is, they proposed an interaction between quantity taken by others and type of others on the amount of consumption. They found an interaction such that consumers

modeled the behavior of a thin confederate more than they modeled the behavior of an obese confederate. They took more candy when she took 30 pieces rather than 2 pieces, but this difference (which might reflect imitation of the model's behavior) was stronger when the model was thin than when she was obese.

Suppose that rather than manipulating the size of the confederate over two levels, McFerran et al. (2010) had a yoked design where pairs of undergraduates participated in the study, and one was cast in the role of confederate and instructed to take 2 or 30 candies, testing the effect on the behavior of the other participant in the pair. Over 100 pairs, suppose they measured the Body Mass Index (BMI) of the 100 confederate models.

How could the authors analyze the interaction and simple effects? The authors could choose to perform a median split and divide individuals into groups with large and small confederate models (or small, medium, and large to allow for nonlinearity). This is not a viable solution because the problems with median splits are well documented: the substantial loss of statistical power from dichotomizing a single predictor variable (e.g., Irwin and McClelland 2001, 2003; Jaccard et al 2006; MacCallum et al. 2002) and the creation of spurious effects when dichotomizing in multiple predictor models (Maxwell and Delaney 1993; Varga et al. 1996). Instead, the authors should use moderated multiple regression and test the model:

(1) $Y = a + bZ + cX + dZX$

where Y is number of candies taken by the participant, X is the BMI of the model, and Z is an indicator variable for number of candies taken by the confederate model. That indicator variable could be dummy coded (0 = 2 candies, 1 = 30 candies) or it could be contrast coded (-1 = 2 candies, +1 = 30 candies).

A significant coefficient *d* in Equation 1 implies that the effect of number of candies taken is moderated by BMI or, equivalently, that the effect of BMI is moderated by number of candies taken. Following detection of a significant interaction, the authors may wish to estimate and test the simple effect of the manipulated variable Z at different levels of X, the BMI of the model. Tests of simple effects of a manipulated or categorical variable at a level of a continuous variable are often called "spotlight" tests: they shine the spotlight on the effect of the manipulated Z at a particular value of X. Spotlight analysis is a technique using basic statistics from regression analysis to analyze the simple effect of one variable at a particular level of another variable, continuous or categorical. This paper is intended to help authors conduct spotlight analyses in various types of experimental and correlational designs and convey their findings more effectively. We show that:

1. Regression terms that authors sometimes interpret as "main effects" are actually simple effects of an interacting variable in a product term (ZX) when other variables in that product (interaction) term are coded as zero.
2. One can shine the spotlight for the simple effect of Z on a particular value of X by adding or subtracting a constant from the original X variable to make the focal value the zero point on the recoded scale.
3. Authors in marketing and allied social sciences have been following a convention of testing simple effects of Z at plus and minus one standard deviation from the mean of X. This one-standard-deviation-from-the-mean spotlight level is arbitrary and hinders generalization across studies.

4. If there are values of X that are particularly meaningful or relevant for theoretical or substantive reasons, simple effects "spotlight" tests should be reported at those values rather than at plus and minus one standard deviation from the mean value of X.
5. If there are no values of X that are particularly meaningful – i.e., all values of X are relevant and interesting values for considering simple effects of the manipulated Z – authors should abandon spotlight tests and report what we call a "floodlight" test of simple effects of Z at all possible values of X. This "floodlight test" from Johnson and Neyman (1936) identifies regions along the X continuum where the simple effect of Z is significant and regions where it is not. It is simple to compute those regions.
6. These same principles can be applied to more complex designs where marketing and consumer researchers have been treading with trepidation. One can readily apply these principles to multiple levels of Z, within participant manipulations of Z, and to higher order factorial designs including one or more measured variables. The principles involve nothing more than basic regression techniques. We discuss certain statistical subtleties in the Appendix and explain the applications to more complex designs in Web Appendix A. Table 1 covers the contents of Web Appendix A.

Insert Table 1 about here

SIMPLE EFFECTS TESTS AND THE MAGIC NUMBER ZERO

Spotlight analysis provides an estimate and statistical test of the simple effect of one variable at specified values of another continuous variable. Aiken and West (1991), Irwin and McClelland (2001), and Jaccard, Turrisi, and Wan (1990) discuss how to conduct spotlight

analyses. We reiterate the key points here to aid understanding of the general principles underlying spotlight analyses, the explanation of specific examples later, and how this relates to our proposed floodlight analysis.

Take the basic moderated multiple regression model shown in Panel A of Table 2 for the hypothetical version of McFerran et al. (2010) describe earlier. The dependent variable (Y), is analyzed as a function of a two-level manipulated variable (Z), a continuous measured variable (X), and their interaction. Z is coded 0 for the group where the model takes 2 candies and 1 for the group where the model takes 30 candies. The model is given by equation 1 above.

Insert Table 2 and Figure 1 about here

In Figure 1A, hypothetical data for such a model were plotted with the continuous variable (X) plotted on the x-axis and two regression lines relating X to the dependent variable Y: one regression line for the Z = 0 group where the model takes 2 candies and one for the Z = 1 group where the model takes 30 candies. The specific estimates are discussed in the next section.

We have noticed that some authors use the continuous value of X in testing the interaction in Equation 1 – i.e, for "analysis" of the interaction. But when it comes time to do simple effects tests to "explicate" the interaction, they revert to using median splits, testing the simple effect of Z at different levels of the now-dichotomized X. This is incorrect. The correct test of simple effects of Z at different levels of X uses the continuous X and spotlight tests.

The simple effect of Z at a given value of X is equivalent to the distance between the regression line for the treatment group and the regression line for the control group. The regression line for the group where the model takes 30 candies, where Z = 1, is found by replacing Z with 1:

(1a) $Y = a + b + cX + dX = (a + b) + (c + d)X$

The intercept, where X = 0, is given by ($a + b$) and the slope is given by ($c + d$). The regression line for the group where the model takes 2 candies, where Z = 0, is found by replacing Z with 0:

(1b) $Y = a + cX$

The intercept, where X = 0, is given by a and the slope is given by c. The simple effect of the manipulation, Z, given by the difference between the lines[1], is therefore:

(1c) $\Delta Y = b + dX$

Equation 1 and Equation 1c make clear that b is the simple effect of Z when X = 0, even though X = 0 may well be outside the range of the data or an impossible value. Equation 1 simplifies to $Y = a + bZ$ where X = 0. Equation 1c–which estimates the simple effect as the difference between two regression lines–simplifies to $\Delta Y = b$ where X = 0.

Zero is a "magic number" in moderated regression. It is "magic" because equation 1 simplifies when either variable has a value of zero. One can add or subtract a constant to a moderating variable to make the coefficients on the other variables reflect simple effects of those variables at particular values of the moderator.

Simple Effect of Categorical Variable Z at a Given Level of Continuous Variable X

Understanding that the coefficient b reflects the simple effect of Z where X = 0 and that the coefficient c reflects the simple effect of X when Z = 0, one can recode X to examine the effect and statistical significance of Z at some value X_{Focal} other than the original X = 0. Simply subtract X_{Focal} from X to create a new variable ($X' = X - X_{Focal}$). Rerun the regression using X' instead of X. The estimate, standard error, and significance test of b' are equivalent to those of $b + dX$ at the focal value because when $X = X_{Focal}$, $X' = 0$. This is shown in Panel B of Table 2.

[1] More generally, the simple effect of Z on Y is given by the derivative of Y with respect to Z.

Simple Slope of Continuous Variable X at a Given Level of Categorical Variable Z

The same "magic number zero" principles can be used if we wish to know the simple effect of the quantitative variable X at a given level of the manipulated Z. We can use the same principle to examine the estimate, standard error, and significance test of the slope of either line. Because the line for the group where the model takes 2 candies, where Z = 0, is given by (1b), the estimate, standard error, and significance test of *c* represent the estimate, standard error, and significance test of the slope of X for that group. To test the slope of X for the group where the model takes 30 candies, simply recode Z such that Z = 0 for the group where the model takes 30 candies and Z = 1 for the 2 candy group. This is shown in Panel C of Table 2.

It is important to note that this is only the case when Z is dummy coded (one group is coded as 0 and the other group is coded as 1). If Z is contrast coded such that one group is coded as -1 and the other is coded as 1, the magic number zero principle still holds such that the coefficient on X still represents the relationship between X and Y when Z = 0, but this no longer represents the simple slope for either group. Instead, if Z is contrast coded, the coefficient on X represents the unweighted average of the two simple slopes (a "main effect" in ANOVA terms). The simple slope for the group represented by Z = -1 is given by (c – d) and the simple slope for the group represented by Z = 1 is given by (c + d).

These two examples, testing both the difference between two regression lines and the slope of a single line by recoding interacting variables, are examples of a broader principle. *In linear models with interaction terms, the estimate, standard error, and significance test of a coefficient on a variable represent the estimate, standard error, and significance test of the simple effect of that variable when all variables with which it interacts are equal to 0.* As noted

by Irwin and McClelland (2001), many scholars incorrectly interpret these parameters as "main effects" rather than as simple effects. This mistake persists in recent marketing research.

Because coefficients *b* and *c* in Equation 1 are interpretable as simple effects when interacting variables are set equal to 0, strategic recoding allows a researcher to examine effect sizes and significance tests at other values of interest[2]. This is not limited to the familiar 2 x Continuous design. We describe other cases later in this paper and in Web Appendix A. In the next section, we will illustrate this point and emphasize the role of focal values in the simple case of two interacting variables in a 2 x Continuous design.

SPOTLIGHT ANALYSIS AT MEANINGFUL FOCAL VALUES

We begin illustrating spotlight analysis in a simple common design. We have two purposes in discussing this design. First, we establish the basic paradigm used in all extensions of the magic number zero in more complex designs. Second, we emphasize the suboptimal nature of the spotlight tests most often reported for these designs in published marketing research, where one examines simple effects of a manipulated variable at plus and minus one standard deviation from the mean of a measured variable.

For this and subsequent examples below, we generated fictitious illustrative data (N = 100) to showcase various analysis methods and results; all of these data were generated to be consistent with plausible predictions made from McFerran et al.'s (2010) results discussed

[2] Rather than recoding X' and redoing the regression analysis for a given X_{Focal}, one can directly estimate the coefficient *b* and its standard error using components of the variance-covariance matrix for the parameters (e.g., Aiken and West 1991, p18). Those adept at using variance-covariance matrices may do so to estimate the standard error. However, others will find recoding and redoing the regression analysis easier, especially as the models become more complex.

above, but no real data were collected for these examples[3]. As before, the dependent variable (Y) is the number of candies taken by the participant. The dichotomous independent variable (Z) is the number of candies taken by the confederate (2 candies, coded as 0 vs. 30 candies, coded as 1). The continuous variable (X) is the BMI of the confederate (M = 21.97, SD = 2.90). We are interested in the effect on quantity taken by the non-confederate participant. We estimate the parameters of the moderated regression model given by equation 1; these are shown in Table 3A and displayed in Figure 1A.

Insert Table 3 about here

The regression results using the untransformed data are not readily interpretable. The significant value of the coefficient d tells us that the interaction is significant. That is, the two experimental groups have different slopes relating BMI of the confederate to number of candies the participant takes. In the 2 candy condition where Z = 0, the slope is c. In the 30 candy condition where Z = 1, the slope is $c + d$. However, because of the magic number zero, a (the intercept) and b (the coefficient for Z) pertain only to when BMI = 0, an impossible value.

Given that we find a significant interaction, at what values of X should we test for simple effects of Z? The convention is to test at one standard deviation above and below the mean, though we argue below that these arbitrary values are not very informative and instead researchers should test at meaningful focal values. There are commonly agreed cutoffs for BMI between underweight and normal weight, normal weight and overweight, and overweight and obese. These cutoffs represent meaningful values, and we would argue that readers should be

[3] The datasets for the 2 x Continuous and 3 x Continuous examples are available in the Web Appendix so that readers may replicate the analyses reported here or do any further examination of these illustrative, hypothetical data.

more interested in knowing the effect of Z at these focal meaningful values that are not sample-dependent than they should be in tests at plus and minus one standard deviation from the sample mean for an idiosyncratic sample. Further, effect sizes for sample-dependent values of X are less likely to generalize than those for meaningful focal values that are the same across studies.

Consequently, we might want to know the effect of Z when X = 25, the cutoff between being normal weight and being overweight; these procedures could be equally applied to any of the other focal cutoffs. To see the effect of choice of large versus small quantity for a borderline overweight confederate, we would simply define a new variable X' = X – 25. We want to set the X value of interest equal to 0—this is the key. We set X' = X – 25 and rerun the resulting model:

$$Y = a' + b'Z + c'X' + d'ZX'$$

Table 3B shows the parameter estimates and tests for an analysis of this model using X' = BMI – 25 instead of X = BMI. It is important to note that 25 is subtracted from raw, not mean-centered, BMI.

The parameters for the model in Table 3B are depicted in Figure 1B. In the figure, a value of X = 25 corresponds to X' = 0. Note that the underlying models in Figures 1A and 1B are identical. In particular, the slopes for the two groups are unchanged by the transformation of X and the estimates and tests are unchanged for the coefficient $c = c'$, the slope for the group observing the confederate take the smaller quantity. Further, there is no change in the interaction coefficient $d = d'$, the difference between the slopes in the two conditions. However, what has changed is that now the coefficient b' ($\neq b$) estimates the difference between the two groups when X = 25, i.e., the effect of the number of candies taken by the observed confederate, estimated for confederates with BMI = 25, the cutoff for being overweight. In this case, there is a significant difference between the two groups when X' = 0 or equivalently, when X = 25. The

intercept *a'* also changes because now it reflects the forecasted value when X' = 0 (i.e., X = 25) and Z = 0; that is, the model predicts that participants who observe a borderline overweight model (BMI = 25) take a small quantity of candies will themselves take about 8.4 candies. In summary, when all terms are included in the model, adding or subtracting a constant to a variable X leaves the underlying model unchanged and only changes the coefficients for the intercept and other variables with which that variable is multiplied. It does not affect the coefficient on terms that include that variable, contrary to what many first expect when learning spotlight tests.

This example illustrates two major points. First by recoding variables in a moderated regression to change what is coded as zero, one can derive simple effect "spotlight" tests for the effect of a variable in the model when other variables are set to zero. Second, there are many cases where authors should break from the convention of conducting spotlight tests at plus and minus one standard deviation from the mean. In fact, we would argue that this convention is almost never the best approach – notwithstanding that we have both used and advocated that approach in prior papers.

There are three main problems of testing at plus and minus one standard deviation. First, if the distribution of the moderator X is skewed, one of those values can be outside the range of the data. Second, if the moderator X is on a coarse scale, it may be impossible to have a value of X exactly equal to plus or minus one standard deviation. Third, if two researchers replicate the same study with samples of very different mean levels of the moderator, it can appear that they fail to replicate each other even when they find exactly the same regression equation in raw score units. This problem is exacerbated by the tendency for authors using the plus and minus one standard deviation approach to fail to report the mean and standard deviation of X. Fernbach et al. (2013) faced these potential problems using Frederick's (2005) CRT scale. The CRT is a

coarse scale (four points ranging from 0 to 3) that can take on skewed distributions that vary substantially across populations. For example, one standard deviation above the mean of Frederick's (2005) MIT sample (M = 2.18, SD = 0.94) would be an impossibly high value; one standard deviation below the mean of his Toledo sample (M = 0.57, SD = 0.87) would be an impossibly low value; and a "low" MIT score would be similar to a "high" Toledo score. Fernbach et al. successfully handled these problems by not using sample-dependent and potentially impossible values of "high" and "low", but rather by testing at the scale endpoints. We expand on these problems and provide further details in Web Appendix B.

In cases where there are meaningful focal values, we recommend a spotlight test, focusing on simple effects of the manipulated variable at judiciously chosen values of the moderator – not at some arbitrary number of standard deviations from the mean. In other cases, no particular value of the moderating variable is particularly focal. In those cases, we will recommend below reporting a "floodlight" analysis of the simple effect of the manipulated variable across the entire range of the moderator, reporting regions where that simple effect is significant. This approach is appropriate when the scale of measurement is "arbitrary" – that is, it is an interval scale of some underlying construct with an unknown zero point.

FLOODLIGHT ANALYSES: SPOTLIGHT ANALYSES FOR ALL VALUES OF X

Spotlight analysis provides a test of the significance of one coefficient at a specific value of another continuous variable, so it is most useful when there is some meaningful value to test. When it is not the case that some values are more meaningful than others or when the researcher anticipates that readers might be interested in other spotlight values, we recommend using an

analysis introduced by Johnson and Neyman (1936) that we dub “floodlight” analysis. Whereas the spotlight illuminates one particular value of X to test, the floodlight illuminates the entire range of X to show where the simple effect is significant and where it is not; the border between these region is known as the Johnson-Neyman point. In essence, this test reveals the results of a spotlight analysis for each and every value of the continuous variable. As Preacher, Curran, and Bauer (2006) note, this eliminates the arbitrariness of choosing high and low values such as one standard deviation above and below the mean.

Johnson and Neyman (1936) introduced the concept and statistical underpinnings for what we are calling “floodlight” analysis. Rogosa (1980, 1981) and Preacher et al. (2006) contributed important later developments. One need not delve into the underlying mathematics to understand the basic concept. The Johnson-Neyman point (or points: there are always two such points that could either straddle a crossover or both be on the same side; see McClelland and Lynch 2012) is simply the value of X at which a spotlight test would reveal a *p*-value of exactly .05 (or whatever alpha one is using). In the case of a 2 x continuous interaction, it is the value of X for which the simple effect of Z is just statistically significant. Values of X on one side of the Johnson-Neyman point yield significant differences between the two groups, while values on the other side do not. In this way, a floodlight shines on the range of values of the continuous predictor X for which the group differences are statistically significant.

Mohr, Lichtenstein, and Janiszewski (2012) provide a recent example of such an analysis and presentation. In their research, they were interested in the effect of the interaction of dietary concern (assessed as a continuous measure averaging items rated on “arbitrary” 1 to 7 scales, yielding, at most, an interval scale of the underlying construct) and health frame on guilt and purchase intention. Because the continuous moderator was assessed on an arbitrary scale without

focal values, they presented the results showing the range over which the simple effect is significant rather than picking sample dependent points without real meaning to the reader.

Conducting a Floodlight Analysis in the 2 x Continuous Case

Floodlight analysis remained obscure for years because of its apparent computational complexity. Now macros for SPSS, SAS, and R (Hayes 2012; Hayes and Matthes 2009; Preacher et al. 2006)[4] make computing Johnson-Neyman points feasible for any researcher. Rather than testing a particular value of X as in spotlight analysis, these macros solve for values of X for which the t-value is exactly equal to the critical value – in other words, values for which a spotlight analysis would give significant results on one side and non-significant results on the other side. Rather than spotlighting a single point, this floodlights the entire range of the data to reveal where differences are and are not significant rather than focusing on one or two arbitrary points. As Potthoff (1964) and Hayes and Matthes (2009) note, these regions do not adjust to account for multiple comparisons across the entire range. However, they do allow one to claim that any spotlight test within that range would be significant.

Researchers preferring not to download and learn new macros can readily perform a floodlight analysis by performing a spotlight analysis for a grid of interesting values, being sure to include the minimum and maximum plausible values of the continuous predictor. Often there is only a discrete set of interesting or plausible values (e.g., points on a 1–to–7 rating scale). For example, Nickerson and colleagues (2003) provide the spotlight values of the coefficient for a list of income ranges that were of interest. If finer precision is desired, it is easy to see between which grid values the spotlight switches from being significant to non-significant; the Johnson-Neyman value must lie within that interval. Iterative spotlight analyses using numbers between

[4] The macros and instructions for using them are available at http://afhayes.com/spss-sas-and-mplus-macros-and-code.html and http://quantpsy.org/interact/index.html.

those two grid values will quickly determine a fairly exact Johnson-Neyman value. Importantly, this iterative process generalizes to more complex designs for which macros often do not exist. The general strategy is to do spotlight analyses on a grid of values for one or more variables and note the regions in which the spotlight values switches from significant to insignificant. Then iterate between those values if more precision is desired.

As an illustration of doing floodlight analysis by an iteration of selected spotlight values, Table 4 displays the difference between the two example groups in number of candies taken (i.e., the vertical distance between the two regression lines in Figure 1) for spotlighted values of BMI between 16 and 32 in steps of 2, along with the associated statistical information for the coefficient describing the difference at each spotlighted value of BMI. The table was constructed by doing nine spotlight regressions. For example, the values in the first row were obtained by computing a new variable X' = X – 16 and then estimating the regression model

$$Y = a' + b'Z + c'X' + d'ZX'$$

The tabled values are the statistical values for the coefficient *b'*.

Insert Table 4 and Figure 2 about here

Note in Table 4 that the difference between groups in number of candies taken switches from being significantly greater than zero for BMI = 26 to not being significantly greater than zero for BMI = 28. Thus, the Johnson-Neyman point must be between BMI = 26 and BMI = 28. Further iteration within the range between 26 and 28 (not presented here) locates the Johnson-Neyman point more precisely at BMI = 27.4. Figure 2A displays the regression lines for both model groups with the filled region (the floodlight) indicating for which values of BMI a spotlight analysis would reveal a significant difference in the number of candies taken between

groups. That is, there is a significant difference between groups for values of BMI between 16 and 27.4, and not a significant difference for BMI values above 27.4. Figure 2B graphs the simple effect of the manipulation as it varies across X, showing that the Johnson-Neyman point is located where the 95% confidence band around the simple effect intersects the X-axis.

To report a floodlight analysis, report the Johnson-Neyman point and range(s) of significance or if using the grid search method, report the range(s) of significance and intervals tested. When graphing the results, show the Johnson-Neyman point or grid points tested and report range(s) of significance. For example, "Regressing candies taken on the manipulation (2 candies = 0, 30 candies = 1), BMI (M = 21.97, SD = 2.90, min = 16.5, max = 29.0), and their interaction revealed a significant interaction ($t(96) = -2.27$, $p < .05$). To decompose this interaction, we used the Johnson-Neyman technique to identify the range(s) of BMI where the simple effect of the manipulation was significant. This analysis revealed that there was a significant positive effect of candies taken by the model on candies taken by the participant for any model BMI less than 27.4 ($B_{JN} = 3.03$, SE = 1.54, p = .05), but not for any model BMI greater than 27.4."

EXAMPLES OF WHEN TO USE SPOTLIGHT AND WHEN TO USE FLOODLIGHT

Sometimes spotlight analysis is more appropriate and sometimes floodlight analysis is more informative. We lay out the relevant considerations here. First, is the scale meaningful with a known correspondence to the underlying construct, or is it arbitrary with an unknown linear mapping from numbers on the scale to levels of the underlying construct? Second, do readers understand certain focal values to have meaningful referents even if not all values have

meaningful referents? Third, are some values of the variable impossible due to coarseness of the scale? See Figure 3 for a simple decision tree to guide the decision of whether to use spotlight or floodlight.

Insert Figure 3 about here

Blanton and Jaccard (2006) decried misuse of "arbitrary metrics" in psychology, wherein values of some interval scale were interpreted as "low" or high." A scale is "arbitrary" when the parameters of the function linking the individual's true score on a latent construct to observed scores is unknown or not transparent. Blanton and Jaccard were particularly critical of the use of IAT scores from the Implicit Association Test. IAT scores are based on reaction times, which as a measure of time have ratio scale properties. But when the reaction time (or difference of reaction times) is interpreted as a measure of latent prejudice, it becomes an arbitrary scale and values of zero no longer are particularly meaningful. (Difference scores computed from interval scale ratings of two objects can be meaningful values if zero truly represents no difference in perceptions/ratings of the objects. See Spiller 2011, Study 4 and Appendix, for an example.)

Most individual difference variables used in marketing and consumer research are arbitrary in that they are interval scales of the underlying constructs with unknown units and origins: propensity to plan, involvement, need for cognition, tightwad-spendthrift, need for uniqueness, etc. We would argue that floodlight tests are likely to be more appropriate than spotlight tests at chosen values of these scales.

Where values are clearly non-arbitrary and focal values are meaningful, spotlight analysis can be particularly illuminating. Consider Study 1 from Leclerc and Little (1997). This provides an early example of clever rescaling of X to generate a meaningful spotlight test to test the effect

of advertising for people who were maximally brand loyal. The authors examined how the effect of advertising content type (Ad Type: picture vs. information) on brand attitude varied as a function of brand loyalty.

Brand loyalty was operationalized as a function of the number of brands purchased in the product category the prior year: fewer brands indicate greater loyalty. Had the authors simply used the raw number of brands, the coefficient on advertising content type would have represented the effect of advertising content type for individuals who purchased zero brands the previous year. This would have been problematic for two reasons. First, non-users who did not purchase any brands the previous year were excluded from the analysis, so this point was outside the range of the data. Second, this was not a substantively meaningful value to test: the theory made predictions regarding brand loyalty, not brand usage.

However, because the theory made predictions for individuals who were brand loyal, it was meaningful to test the simple effect for individuals who purchased a single brand the previous year (i.e., those who were completely brand-loyal). Thus, the authors created a new variable, Switching, calculated as number of brands purchased minus one. The authors regressed Brand Attitude on Ad Type, Switching, and Ad Type * Switching. The simple effect of advertising content type could therefore be interpreted as the simple effect when Switching was equal to zero; in other words, the simple effect for brand loyalists. Transforming one variable such that zero took on a substantively meaningful value provided the reader with information about an easily interpretable simple effect. Spotlight was particularly useful in this case; moreover, because loyalty can take on only integer values, it would be better to present spotlights at any integer values likely to be of interest to readers rather than arbitrary and impossible values one standard deviation above and below the mean.

The same point applies to our previous hypothetical extension of McFerran et al. (2010). Had relative weight been measured using a subjective 1 to 7 scale rather than BMI, we would advocate using a floodlight analysis to consider ranges of significance rather than meaningless values one standard deviation above and below the mean.

Interval scales can become nonarbitrary when researchers develop norms for where a particular score in the distribution lies across some reasonably representative sample in a population of consumers. Churchill (1979) advocated this development of norms as the last step of scale development. But this step of norming has not been a part of practice in most marketing and consumer research on scale development, including our own.

We argue that in cases like our McFerran et al. (2010) example, it is more meaningful to report spotlight tests at specific values of an interval or ratio scale – which facilitates comparisons across studies using the same scale with samples drawn from different populations – than to bury the metric of the original scale by reporting plus and minus some sample-dependent standard deviation. This allows accumulation of findings over time about the range of values of the scale where the simple effect of some independent variable is substantively and statistically significant. Edwards and Berry (2010) argue that moving beyond hypotheses that merely postulate the sign of some effect to specifying range of values where the effect holds can increase theoretical precision.

MAGIC NUMBER ZERO FOR SPOTLIGHTS IN OTHER COMMON DESIGNS

Thus far we have discussed spotlight and floodlight analyses in the simple 2 x continuous case. In the next section, we show the reader a simple, easy-to-implement method for

accomplishing spotlight analysis in other common designs in addition to the standard 2 x Continuous case discussed above. It should be apparent that floodlight tests can be derived in these designs too by the iterative grid approach like that in Table 4.

In reading the literature, it is clear that when designs vary from the standard 2 x Continuous design, authors take a variety of inappropriate strategies to analyze simple effects, including median-splitting one or more continuous variables, breaking down a higher-order interaction into separate subsamples and running piecewise analyses on each, and misinterpreting simple effects at one level of a variable as main effects across all levels of a variable. The cases below provide a better way of conducting such analyses.

Below we show how to extend these principles to a) the case of a 2 x Continuous design when Z is manipulated within participants, and b) 2 x 2 x Continuous design when all factors are between participants. Web Appendix A extends these principles to two other cases, the 3 x Continuous design and the case in which X and Z are both continuous. We also consider models with quadratic terms. We emphasize that by no means are these the only designs for which spotlight can be used. Instead, these are representative examples of common designs, and the basic principle used in these designs (recognition of "the magic number zero") can be applied to *every other design* that uses a linear model (including logistic regression) for analysis. For each design, we build on the basic extension of McFerran et al. (2010) described above. Analysis templates for Cases 1, 2, and 3 are given in Web Appendix A along with the extensions above.

Case #1: 2 x Continuous when the Manipulation of Z is Within Subjects

Imagine a version of our original example with two levels of the manipulated factor Z (0 = confederate took 2 candies, 1 = confederate took 30 candies) and a continuous measure of X = BMI. This time, however, let Z be a repeated measures factor. In this case, one simply creates a

contrast score for each subject showing the effect of the manipulation for that subject: $Z_{contrast} = Y_{30} - Y_2$ (see Judd, McClelland, and Ryan 2009 or Keppel and Wickens 2004); one could similarly create contrast scores for within-subject designs with more than 2 levels. One then analyzes the $Z_{contrast}$ scores as a function of X = BMI:

(3) $Z_{contrast} = a + bX$

Extending the principle of the magic number zero, the test of the intercept, *a*, in this analysis is the predicted $Z_{contrast}$ score when X = 0. The coefficient *b* now is equivalent to a test of the interaction of X with Z in the original design. To create a spotlight test of the effect of the repeated factor Z at the borderline between normal and overweight, create X' = X – 25. Rerun the regression $Z_{contrast} = a' + b'X'$. Now the test of the intercept *a'* is the effect of the repeated factor Z at the new zero point associated with the chosen level of X.

Case #2: 2 x 2 x Continuous

Often, one may be interested in how a continuous variable moderates a 2 x 2 interaction, resulting in a three-way interaction. For example, in addition to manipulating quantity taken, we might also manipulate the perceived health of the item being considered (candy vs. granola, as in Study 1 of McFerran et al. 2010). The prediction might be that attenuation of assimilation only occurs for unhealthy food because participants are cued to be more vigilant when food is unhealthy than when it is healthy. (McFerran et al. found this not to be the case.) The model for this design is:

(4) $Y = a + bZ + cW + dX + eZW + fZX + gWX + hZWX$

Z and X are coded the same as they were in the opening example, and W is coded 0 for candy and 1 for granola. If the parameter *h* testing the three-way interaction is significant, it becomes relevant to test the simple interaction of two of the variables at different levels of the

third variable. The coefficient *e* tests the simple ZW interaction when X = 0. (It does *not* test the ZW interaction that would be evident in plotting the ZW cell means, collapsing over levels of X.) The coefficient *f* tests the simple ZX interaction when W = 0. The coefficient *g* tests the simple WX interaction when Z = 0. To follow up a simple two-way interaction, one tests the simple-simple effect of one of the variables holding constant the other two. In this model, *b* represents the simple-simple effect of Z when W = 0 and X = 0; *c* represents the simple-simple effect of W when X = 0 and Z = 0; and *d* represents the simple-simple effect of X when Z = 0 and W = 0.

Zero is a magic number in this analysis too. Every coefficient is interpreted as the effect of that variable (or interaction) when all variables with which that term interacts are set to 0, causing them to drop out of the model.

Suppose that we obtained a significant three-way interaction ZWX and wanted to follow up with tests of the simple ZW interaction at meaningful levels of X. We would recode X' = X – 25 just as in each of the previous examples. When X' = 0, the new coefficient on ZW, *e'*, represents the simple interaction between quantity taken and type of snack taken (granola v. candy) when X' = 0, which corresponds to a BMI of 25.

Spotlight analysis in a 2 x 2 x Continuous design requires application of the exact same principle used in the previous cases: recoding variables such that 0 represents the value of a variable at which you are interested in the simple effect of the other variables.

Web Appendix A includes detailed explanations of how to do simple effects tests in Cases #1 and #2 above. It also covers three additional cases (Case #3, 3 x Continuous, Case #4: Continuous x Continuous, and Case #5: Quadratic). It should be evident for each of these cases that one can easily accomplish floodlight analyses by iterating to redefine the value of X when X' = 0 as in Table 4 within that particular design. Of course, one can extend these analyses to

other designs not described here; for example, one could examine 2 x 3 x Continuous by combining the strategies used in the 3 x Continuous and 2 x 2 x Continuous cases.

CONCLUSION

"Spotlight" tests reflect the simple effect of a variable Z at different levels of an interacting variable X. These tests were popularized by Aiken and West (1991), Jaccard et al. (1990), and Irwin and McClelland (2001), but they remain misunderstood. We have shown that these tests rely on basic multiple regression, simply changing the coding of variables to alter the zero point to give direct tests of interest from the parameters of a moderated regression model.

Some researchers, reviewers and editors seem wary or uncertain of using these tests in anything but the simple case of a dichotomous manipulated variable Z and a continuous measured variable X. For that reason, they fall back on the flawed practice of dichotomizing continuous variables when faced with more complex designs, or they use the continuous variable to test the significance of the interaction but dichotomize to graph the interaction or do simple effects tests. We show how the general principle of "the magic number zero" can be applied to derive ready tests of simple interactions and simple-simple effects in an array of more complex designs. When interacting variables are coded such that zero represents focal values, those interacting variables drop out of the model at their focal values. The remaining terms in the model are then interpreted as simple effects at those focal values of interacting variables.

We criticize the common practice of reporting spotlight tests of the simple effect of Z at plus and minus one standard deviation from the mean on an interacting variable X. We argue that this is almost never optimal because those tests and estimates are sample-dependent in defining

high and low values of X, because it is possible that these refer to impossible values of X, and because readers are not inherently more interested in the effect of Z at plus and minus one standard deviation than at values of X somewhat higher or lower. We argue that in come cases, researchers overlook that there may be "focal" values that are of particular interest and we encourage use of these more judiciously chosen levels of the continuous variable for spotlight tests.

There are many cases where researchers apply spotlight analysis where no particular value of the continuous variable is focal. In these cases, we recommend abandoning the convention of testing spotlights at plus and minus one standard deviation from the mean. Instead, we recommend use of a related test that shows ranges of the continuous variable where the simple effect of a second variable is significant and where it is not. Johnson and Neyman (1936) originally reported this technique. We dub this a "floodlight" analysis, as it illuminates the entire range of the data rather than spotlighting a single point. Reporting floodlight analyses allows readers an efficient way to infer whether two groups differ at any given point of interest and facilitates comparing and integrating findings across multiple samples with different sample distributions.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park, CA: Sage Publications.

Belia, Sarah, Fiona Fidler, Jennifer Williams, and Geoff Cumming, G. (2005) "Researchers Misunderstand Confidence Intervals and Standard Error Bars. *Psychological Methods,10(4),* 389-396.

Blanton, Hart, and James Jaccard (2006), "Arbitrary Metrics in Psychology," *American Psychologist*, 61 (1), 27–41.

Cohen, Jacob, and Patricia Cohen (1983). *Applied multiple regression/correlation analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Edwards, Jeffrey R., and Berry, James W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13 (4), 668-689.

Fernbach, Philip M., Steven A. Sloman, Robert St. Louis, and Julia N. Schube (2013), "Explanation Fiends and Foes: How Mechanistic Detail Determines Understanding and Preference." *Journal of Consumer Research*, forthcoming.

Frederick, Shane (2005), "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19 (4), 25-42.

Hayes, Andrew. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/ public/process2012.pdf

Hayes, Andrew F., and Jörg Matthes (2009), "Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations," *Behavior Research Methods*, *41* (3) 924-36.

Irwin, Julie R. and Gary H. McClelland (2001), "Misleading Heuristics and Moderated Multiple Regression Models," *Journal of Marketing Research*, 38 (1), 100-109.

------ and ------ (2003), "Negative Effects of Dichotomizing Continuous Predictor Variables," *Journal of Marketing Research*, 40 (August), 366-371.

Jaccard, James, Vincent Guilamo-Ramos, Margaret Johansson and Alida Bouris (2006): Multiple Regression Analyses in Clinical Child and Adolescent Psychology, Journal of Clinical Child & Adolescent Psychology, 35:3, 456-479.

------, Robert Turrisi, and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*. Newbury Park: Sage.

Johnson, Palmer O. and Leo C. Fay (1950), "The Johnson-Neyman Technique: Its Theory and Application," *Psychometrika*, 15 (4), 349-367.

------ and Jerzy Neyman. (1936). "Tests of certain linear hypotheses and their application to some educational problems." *Statistical Research Memoirs*, 1, 57-93.

Judd, Charles M., Gary H. McClelland, and Carey S. Ryan (2009), *Data Analysis: A Model Comparison Approach (Second Edition)*. New York: Routledge.

Keppel, Geoffrey and Thomas D. Wickens (2004), *Design and Analysis: A Researcher's Handbook*. New York: Pearson.

Leclerc, France, and John D. C. Little (1997), "Can Advertising Copy Make FSI Coupons More Effective?" *Journal of Consumer Research*, 34 (4), 473-84.

MacCallum, Robert C., Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker (2002), "On the practice of dichotomization of quantitative variables," *Psychological Methods*, 7(1), 19-40.

Maxwell, Scott E. and Harold D. Delaney (1993), "Bivariate median splits and spurious statistical significance," *Psychological Bulletin*, 113(1), 181-190.

McClelland, Gary, and John G. Lynch, Jr. (2011), "Power Considerations in Simple Effects Tests in Moderated Regression." Unpublished working paper, University of Colorado-Boulder.

McFerran, Brent, Darren W.Dahl, Gavan J. Fitzsimons, and Andrea C. Morales (2010), "I'll Have What She's Having: Effects of Social Influence and Body Type on the Food Choices of Others," *Journal of Consumer Research*, 36 (April), 915-929.

Mohr, Gina S., Donald R. Lichtenstein, and Chris Janiszewski (2012), "The Effect of Marketer-Suggested Serving Size on Consumer Responses: The Unintended Consequences of Consumer Attention to Calorie Information," *Journal of Marketing*, 76 (1), 59-75.

Nickerson, Carol, Norbert Schwarz, Ed Diener, and Daniel Kahneman (2003), "Zeroing in on the Dark Side of the American Dream: A Closer Look at the Negative Consequences of the Goal for Financial Success. *Psychological Science, 14(6),* 531–536.

Potthoff, Richard F. (1964), "On the Johnson-Neyman Technique and Some Extensions Thereof," *Psychometrika*, 29 (3), 241-56.

Preacher, Kristopher J., Patrick J. Curran, and Daniel J. Bauer (2006), "Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis," *Journal of Educational and Behavioral Statistics*, 31 (3), 437–448.

Rogosa, David (1980), "Comparing Nonparallel Regression Lines," *Psychological Bulletin*, 88 (2), 307-321.

_____ (1981), "On the Relation Between the Johnson-Neyman Region of Significance and the Statistical Test of Nonparallel Within-Group Regressions," *Educational and Psychological Measurement*, 41 (1), 73-84.

Schenker, Nathaniel and Jane F. Gentleman (2001), "On Judging the Significance of Differences by Examining the Overlap between Confidence Intervals," *The American Statistician, 55(3)*, 182-186.

Spiller, Stephen A. (2011), "Opportunity Cost Consideration," *Journal of Consumer Research*, 38 (December).

Case Number	Design	Covered
0 (Base Case)	2 x Continuous	Main Text p6 and Table 2, Web Appendix A Table W1
1	2 (Within) x Continuous	Main Text p22, Web Appendix A p41 and Table W2
2	2 x 2 x Continuous	Main Text p23, Web Appendix A p42 and Table W3
3	3 x Continuous	Web Appendix A p44 and Table W5
4	Continuous x Continuous	Web Appendix A p46
5	Quadratic	Web Appendix A p47

Table 1. Index of where to find building-block designs for spotlight and floodlight analyses.

SIMPLE EFFECTS IN A 2 x CONTINUOUS DESIGN

$$Y = a + bZ + cX + dZX$$

A. BASELINE ANALYSIS

	Intercept	Manipulation	Measured Variable	Manipulation * Measured
		Z	X	ZX
Coding		0 = Control 1 = Treatment	Raw scale	
Coefficient	*a*	*b*	*c*	*d*
Interpretation	Estimate of Y when Z = 0 and X = 0, i.e., for Control group when X = 0	Simple effect of Treatment vs. Control vs. Control when X = 0	Simple slope of Measured Variable on Y when Z = 0, i.e., for Control group	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

B. TEST THE SIMPLE EFFECT OF TREATMENT VS. CONTROL AT FOCAL VALUE $X = X_{Focal}$
BY RECODING X SO THAT IT DROPS OUT OF THE EQUATION

		Z	X'	ZX'
Coding		0 = Control 1 = Treatment	$X' = X - X_{Focal}$	
Coefficient	*a'*	*b'*	*c'*	*d'*
Equivalent to	$a + cX_{Focal}$	$b + dX_{Focal}$	*c*	*d*
Interpretation	Estimate of Y when Z = 0 and X' = 0, i.e., for Control group when $X = X_{Focal}$	Simple effect of Treatment vs. Control when X' = 0, i.e., when $X = X_{Focal}$	Simple slope of Measured Variable on Y when Z = 0, i.e., for Control group	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

C. TEST THE SIMPLE SLOPE OF X IN TREATMENT GROUP BY RECODING Z SO THAT IT DROPS OUT OF THE EQUATION

		Z"	X	Z"X
Coding		1 = Control 0 = Treatment	Raw scale	
Coefficient	*a"*	*b"*	*c"*	*d"*
Equivalent to	*a + b*	*-b*	*c + d*	*-d*
Interpretation	Estimate of Y when Z" = 0 and X = 0, i.e., for Treatment group when X = 0	Simple effect of Treatment vs. Control when X = 0	Simple slope of Measured Variable on Y when Z" = 0, i.e., for Treatment group	Difference in slope of Measured Variable between Control (Z" = 1) and Treatment (Z" = 0)

Table 2. Simple effects in a 2 x Continuous design.

A.

Variable	Coefficient	Estimate	Standard Error	*t*	*p*
intercept	*a*	3.65	3.93	0.93	.356
coded number of candies taken by confederate	*b*	18.56	5.54	3.35	.001
confederate BMI	*c*	0.19	0.17	1.10	.273
coded number of candies taken by confederate * confederate BMI	*d*	-0.57	0.25	-2.27	.026

B.

Variable	Coefficient	Estimate	Standard Error	*t*	*p*
intercept	*a'*	8.44	0.68	12.39	<.001
coded number of candies taken by confederate	*b'*	4.39	1.05	4.18	<.001
confederate BMI - 25	*c'*	0.19	0.17	1.10	.273
coded number of candies taken by confederate * (confederate BMI – 25)	*d'*	-0.57	0.25	-2.27	.026

Table 3. Regression results of fictitious illustrative data. Panel A: Results in raw metric (X = BMI). Panel B: Results after transformation (X' = BMI – 25) to examine the simple effect for individuals with a BMI of 25.

BMI	Group Difference (Candies Taken)	Lower 95% CI	Upper 95% CI	*t*(96)	*p*
16	9.49	6.20	12.80	5.73	<0.0001
18	8.36	5.92	10.8	6.82	<0.0001
20	7.22	5.49	8.95	8.28	<0.0001
22	6.09	4.65	7.52	8.43	<0.0001
24	4.95	3.20	6.71	5.60	<0.0001
26	3.82	1.35	6.29	3.07	0.003
28	1.55	-0.64	6.01	1.60	0.11
30	1.55	-2.69	5.80	0.72	0.47
32	0.42	-4.77	5.61	0.16	0.87

Table 4. Spotlight analyses of the difference between groups in number of candies taken for a systematic selection of BMI values between the minimum and maximum.

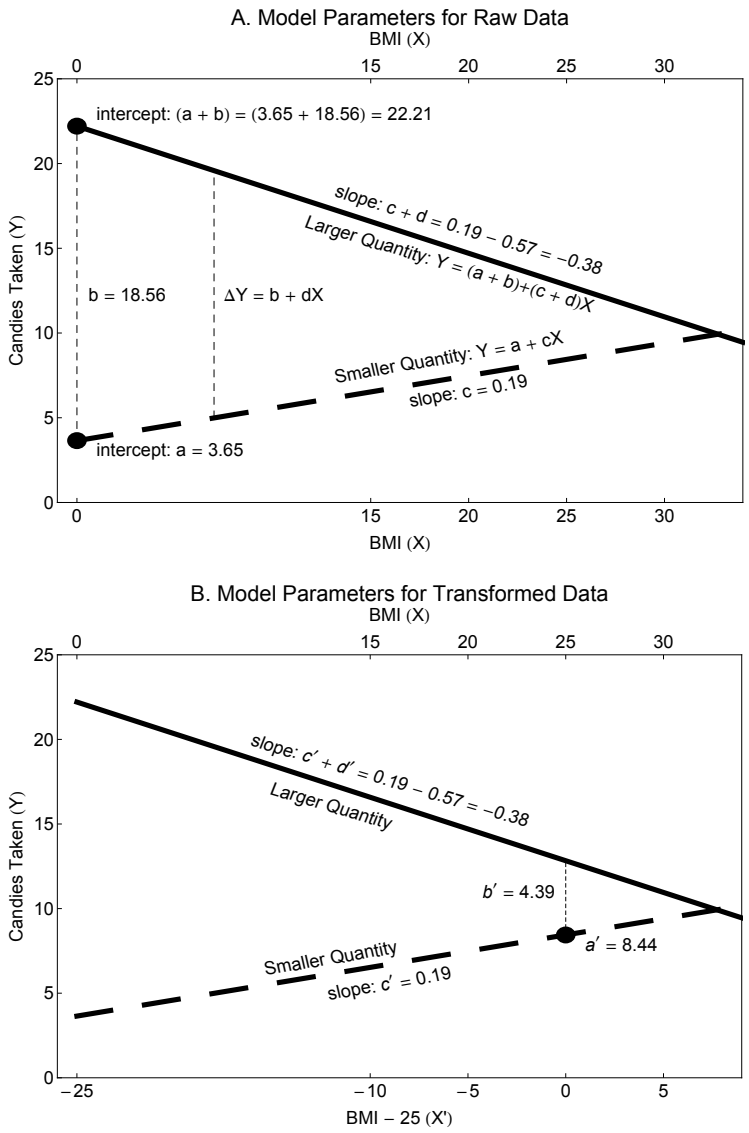

Figure 1. Graphical interpretation of regression parameters from Table 2. The dashed lines represent the smaller quantity group where Z = 0; the solid lines represent the larger quantity group where Z = 1. Panel A shows the regression results using the untransformed data. The coefficient on quantity, *b*, reflects the effect of quantity for a BMI of 0, an impossible value that lies well outside the range of the data. Panel B shows the regression results using the transformed data, recoded such that the definition of borderline overweight (a BMI of 25) lies at 0. Everything about the graph is exactly the same, other than the recoded X-axis. The statistical test still is a test at 0, but now 0 corresponds to a substantively meaningful value.

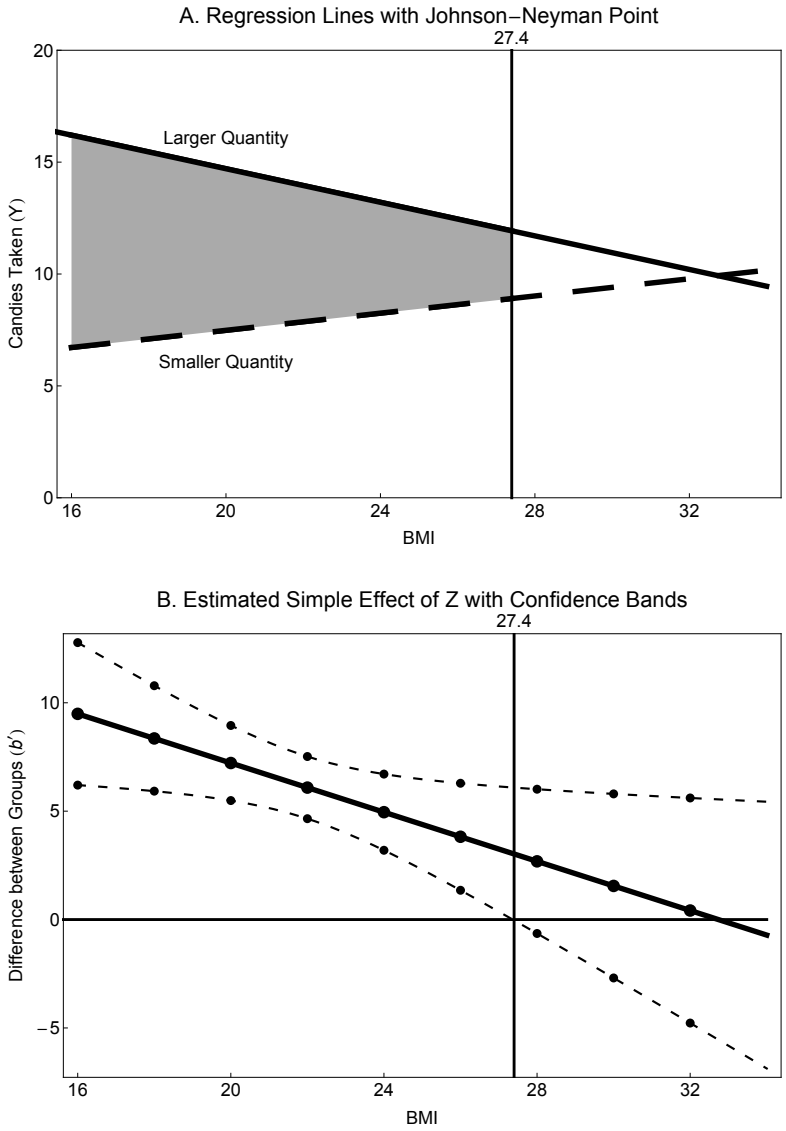

Figure 2. Panel A shows a floodlight of the region of BMI values (filled area below 27.4) for which a spotlight test would reveal significant differences between the two model groups. Panel B shows a graph of the estimated simple effect (the distance between the two regression lines in Panel A) with confidence bands. Confidence bands are narrowest at mean BMI (M = 21.97). The Johnson-Neyman point in Panel A aligns with the intersection of the confidence band and the X-axis in Panel B. The crossover point in Panel A aligns with the intersection of the estimated simple effect and the X-axis in Panel B.

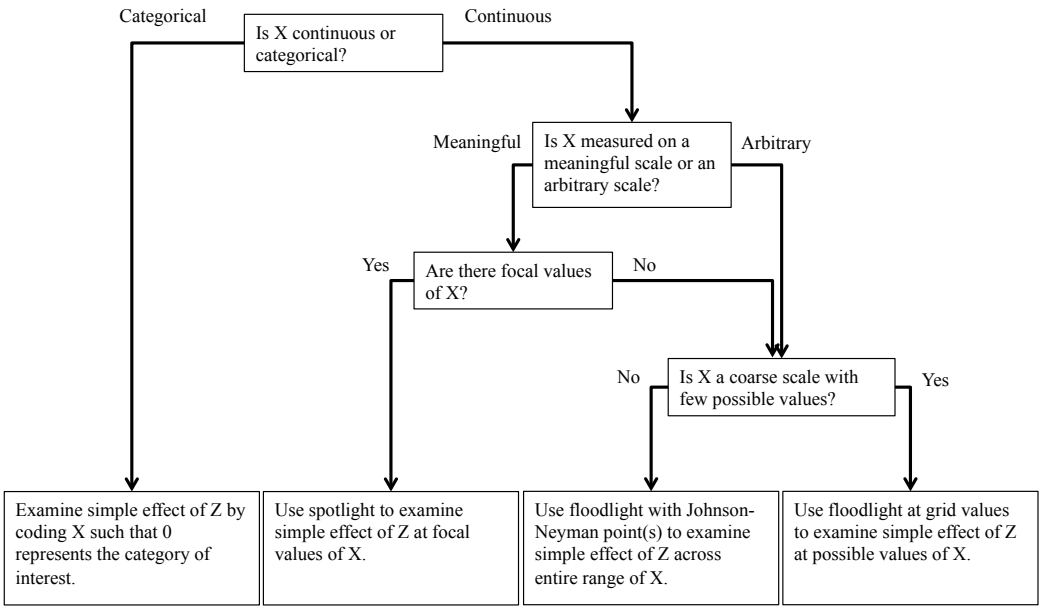

Figure 3. Floodlight decision tree. This decision tree guides the decision of when to use spotlight analysis and when to use floodlight analysis to examine the simple effect of Z across a moderating variable X in a model of the form $Y = a + bZ + cX + dZX$.

APPENDIX: STATISTICAL AND POWER CONSIDERATIONS IN SPOTLIGHT AND FLOODLIGHT TESTS

This section summarizes statistical parameter estimation and power considerations in spotlight and floodlight analysis in a design with a manipulated Z with two levels (dummy coded 0 = Control, 1 = Treatment) and X coded as a continuous variable in its raw metric, analyzed via Equation 1.

(1) $Y = a + bZ + cX + dZX$

1. Power to detect the simple effect of Z varies with X. This is true because both the numerator and the denominator of the F test for the simple effect of Z change with X. A nonzero interaction of X and Z implies that there is some value of X where the regression lines for the two levels of Z cross over, although this crossover may occur outside the range of one's data. Johnson and Neyman (1936) proved that one can find two values of X where the effect of Z is exactly significant. McClelland and Lynch (2012) demonstrated that it is possible to have real data where the effect of Z is significant to the right or left of the crossover point of the interaction, but not significant as one moves further away the crossover. To guard against this, be sure to conduct a spotlight test at minimum and maximum values of X.
2. In the main text we discussed how researchers replicating the same experiment with a manipulated Z and a measured X might find exactly the same regression equation but perceive that they had failed to replicate each other's findings if the two studies used samples with high versus low average values of the measured variable X. Now consider that these two exact replicates were analyzed using spotlights at the same focal value of the moderator, X_{Focal} as expressed in raw score units. The statistical tests on the simple effect of the

manipulated variable Z at X_{Focal} will not match in the two replicates, because the standard error of the coefficient is smaller when the focal value F is closer to the sample mean (McClelland and Lynch 2012).

3. Some methodologically sophisticated colleagues have expressed skepticism when told that the simple effect of a manipulated variable Z is significant at a value of X two standard deviations from the mean. Their statistical intuition is that the analysis relies upon a small subset of cases far from the mean. What they are missing is that the solution to the moderated regression uses all of the data from the study. Like any regression, the spotlight statistical tests of the simple effect of Z at a given level of X reflect the fact that we have wider confidence intervals for the predicted value of Y when X is far from the mean of the data than when it is close to the mean.
4. The use of spotlight tests at each value of an arbitrary but coarse scale is different from treating each value as discrete levels of a categorical factor in an ANOVA. In that latter case, the power of the test of the simple effect of a manipulated variable at a level of the moderator variable is affected only by the number of cases at that level of the moderator variable. In contrast, the spotlight test is a regression parameter estimate that treats the moderator as a continuous variable. In this case, the power of the spotlight test of the simple effect of the manipulated factor is affected by the entire dataset including all possible values of the moderator. In this case, there is no special danger of testing for extreme values of the moderator that are inside the range of the data. Like any regression, the spotlight statistical tests of the simple effect of Z at a given level of X reflect the fact that we have wider confidence intervals for the predicted value of Y when X is far from the mean of the data than when it is close to the mean.

5. Even when the true interaction has a coefficient *d* exactly equal to zero, merely *including* the interaction term also causes the standard error of *b* to change with a rescaling of X, because now *b* is explicitly testing the effect of Z at the value of X coded as 0. For example, consider a model where X takes on values from 1 to 7 with a mean of 4 and Z is dummy coded. Assume that the true interaction is zero, and when one estimates Equation 1, the coefficient *d* on the interaction is exactly zero. The standard error on the coefficient b is the standard error when X = 0, outside the range of the data. One will get a smaller estimate of the standard error and a larger t test on the parameter *b* if one mean centers X' = X – 4, so that now 0 is coded at the mean. The standard error of *b*, the distance between the two lines, is smallest at the mean of X; as one moves farther away from the mean of X, the standard error increases. Thus, even if the estimate of the interaction is exactly equal to 0, the test of the simple effect of Z will be less powerful if tested far from the mean than if tested at the mean. This has practical implications because if X ranges from 1 to 7 and the effects of Z, X, and ZX on Y are modeled using equation (1), even if the effects of X and ZX are exactly 0, the significance test of Z will be misleading or at least misunderstood if X is analyzed in its raw metric. Note that this effect of rescaling X on the estimate and standard error of the effect of Z does *not* occur if no interaction term is included in the model – i.e, for a "main effects" only model $Y = a + bZ + cX$. In that model, neither coefficient *b* nor *c* changes, nor do the standard errors on *b* and *c* change when a constant is added or subtracted from Z or X, though the estimate of *a* changes.

Web Appendix A

This web appendix is intended to serve as a standalone reference guide for conducting spotlight and floodlight analyses after one has read the main text. These materials repeat and expand upon the two cases discussed in the text and add three additional cases. These building blocks can be combined to examine spotlight or floodlight analyses of any linear model. An analysis template for the base case (Case #0: 2 x Continuous) is given in Table W1.

Case #1: 2 x Continuous when the Manipulation of Z is Within Subjects

Imagine a version of our original example with two levels of the manipulated factor Z (0 = confederate took 2 candies, 1 = confederate took 30 candies,) and a continuous measure of X = BMI. This time, however, let Z be a repeated measures factor. In this case, one simply creates a contrast score for each subject showing the effect of the manipulation for that subject: $Z_{contrast} = Y_{30} - Y_2$ (see Judd, McClelland, and Ryan 2009 or Keppel and Wickens 2004); one could similarly create contrast scores for within-subject designs with more than 2 levels. One then analyzes the $Z_{contrast}$ scores as a function of X = BMI:

(W1) $Z_{contrast} = a + bX$

Extending the principle of the magic number zero, the test of the intercept, *a*, in this analysis is the predicted $Z_{contrast}$ score when X = 0. The coefficient *b* now is equivalent to a test of the interaction of X with Z in the original design. To create a spotlight test of the effect of the repeated factor Z at the borderline between normal and overweight, create X' = X – 25. Rerun the regression $Z_{contrast} = a' + b'X'$. Now the test of the intercept *a'* is the effect of the repeated factor Z at the new zero point associated with the chosen level of X. An analysis template for Case 1 is given in Table W2.

Case #2: 2 x 2 x Continuous

Often, one may be interested in how a continuous variable moderates a 2 x 2 interaction, resulting in a three-way interaction. For example, in addition to manipulating quantity taken, we might also manipulate the perceived health of the item being considered (candy vs. granola, as in Study 1 of McFerran et al. 2010), with all factors manipulated between-subjects. The prediction might be that attenuation of assimilation only occurs for unhealthy food because participants are cued to be more vigilant when food is unhealthy than when it is healthy. (McFerran et al. found this not to be the case.) The model for this design is:

(W2) $Y = a + bZ + cW + dX + eZW + fZX + gWX + hZWX$

Z and X are coded the same as they were in the opening example, and W is coded 0 for candy and 1 for granola. If the parameter *h* testing the three-way interaction is significant, it becomes relevant to test the simple interaction of two of the variables at different levels of the third variable. The coefficient *e* tests the simple ZW interaction when X = 0. (It does *not* test the ZW interaction that would be evident in plotting the ZW cell means, collapsing over levels of X.) The coefficient *f* tests the simple ZX interaction when W = 0. The coefficient *g* tests the simple WX interaction when Z = 0. To follow up a simple two-way interaction, one tests the simple-simple effect of one of the variables holding constant the other two. In this model, *b* represents the simple-simple effect of Z when W = 0 and X = 0; *c* represents the simple-simple effect of W when X = 0 and Z = 0; and *d* represents the simple-simple effect of X when Z = 0 and W = 0.

Zero is a magic number in this analysis too. Every coefficient is interpreted as the effect of that variable (or interaction) when all variables with which that term interacts are set to 0 because they drop out of the model. Suppose that we obtained a significant three-way interaction ZWX and wanted to follow up with tests of the simple ZW interaction at meaningful levels of X.

We would recode X' = X – 25 just as in each of the previous examples. When X' = 0, the new coefficient on ZW, *e'*, represents the simple interaction between quantity taken and perceived healthiness when X' = 0 which corresponds to a BMI of 25.

Spotlight analysis in a 2 x 2 x Continuous design requires application of the exact same principle used in the previous cases: recoding variables such that 0 represents the value of a variable at which you are interested in the simple effect of the other variables. An analysis plan for Case 2 is given in Table W3.

If one is interested in testing a 2 x 2 x Continuous design where the second factor, W, is manipulated within-subject, some effects are within-subject effects and some effects are between-subject effects. To test these, one can combine the strategies from Case #1 and Case #2. To examine the between-subject effects ("main effects" of Z, X, and the ZX interaction), calculate $W_{Average}$ as in W3 and regress it on Z, X, and ZX as in W4.

(W3) $W_{Average} = (Y_{W=1} + Y_{W=0}) / 2$

(W4) $W_{Average} = a + bZ + cX + dZX$

a represents the estimate of Y where Z = 0 and X = 0, averaged across levels of W. *b* represents the simple effect of Z on Y at X = 0, averaged across levels of W. *c* represents the simple effect of X on Y at Z = 0, averaged across levels of W. *d* represents how the effect of Z on Y changes with X, averaged across levels of W. One could similarly calculate $Z_{Average}$ in Case #1, though it is unlikely to be a useful metric if one is interested in the within-subject manipulation.

To examine the within-subject effects (all terms involving W), calculate $W_{Contrast}$ as in W5 and regress it on Z, X, and ZX as in W6.

(W5) $W_{Contrast} = Y_{W=1} - Y_{W=0}$

(W6) $W_{Contrast} = a' + b'Z + c'X + d'ZX$

a' represents the simple simple effect of W at Z = 0 and X = 0. *b*' represents the simple interaction of Z with W on Y at X = 0. *c*' represents the simple interaction of X with W on Y at Z = 0. *d*' represents the three-way interaction of Z, X, and W on Y.

Case #3: 3 x Continuous

Imagine that we are replicating the original 2 x Continuous example with a hypothetical variant of McFerran et al. (2010). Instead of manipulating Z at two levels, we add a third condition in which the confederate stands next to the candy but never has an opportunity to take any candy. Now there are three between-subjects conditions: take a large quantity; take a small quantity; no opportunity to take any quantity. Will participants be affected by the mere presence of another when that confederate does not have an opportunity to make a choice? This requires a slightly different analysis plan, but uses the same basic principle of the magic number zero. Of course, one needs *k*-1 dummy variables (or contrast coded variables) to represent *k* levels of a categorical variable. With a 3-level variable, two dummy variables are required. For our example, Z_1 is coded 1 for small quantity, 0 for large quantity or no opportunity, and Z_2 is coded 1 for large quantity, 0 for small quantity or no opportunity. We also need two variables to represent the 2 degrees of freedom interactions, Z_1X and Z_2X. The base model for this design is:

(W7) $Y = a + b_1Z_1 + b_2Z_2 + cX + d_1Z_1X + d_2Z_2X$

This equation can be rewritten as:

(W7a) $Y = (a + cX) + (b_1 + d_1X)Z_1 + (b_2 + d_2X)Z_2$

Written this way, it becomes clear that the effect of Z_1 is $(b_1 + d_1X)$, so b_1 represents the effect of Z_1 (i.e., the difference between the small quantity and no opportunity conditions) when X is equal to 0; similarly, the effect of Z_2 is $(b_2 + d_2X)$, so b_2 represents the effect of Z_2 (i.e., the

difference between the large quantity and no opportunity conditions) when X is equal to 0. Interpreting X also requires great care. The equation can be rewritten again as:

(W7b) $Y = (a + b_1Z_1 + b_2Z_2) + (c + d_1Z_1 + d_2Z_2)X$

Writing it in this form makes it clear that the effect of X is $(c + d_1Z_1 + d_2Z_2)$, so c represents the effect of X when Z_1 and Z_2 are both equal to 0. In the current example, this means that c represents the simple slope of X for the no opportunity condition. Table W4A presents the statistical analysis of illustrative fictitious data ($N = 150$; $M_{BMI} = 21.81$, $SD_{BMI} = 2.91$). These are the same data for the small and large quantity groups presented in the base case, with a third group of observations added for the condition in which there is no opportunity for the confederate to take any candy. With this coding, the slope for X (BMI) pertains only to the no opportunity group, revealing a flat line ($c = 0.02$, $t(144) = 0.12$, $p = .90$).

To understand the effect of small quantity vs. no opportunity or large quantity vs. no opportunity, it is important to recode X—otherwise b_1 and b_2 represent the effects for a confederate with a BMI of 0. Setting X' = X – 25 and examining the coefficients on Z_1 and Z_2, we can consider the effects of confederates who are borderline overweight. Table W4B presents the results after the recoding of X. Estimated for confederates with BMI = 25 (i.e., X' = 0), taking a larger quantity of candies significantly increased the number of candies taken by the observer ($b_2' = 3.12$, $t(144) = 2.98$, $p = .003$) relative to the no opportunity condition, whereas taking a smaller quantity of candies did not significantly decrease the number of candies taken by the observer relative to the no opportunity condition ($b_1' = -1.27$, $t(144) = -1.30$, $p = 0.19$).

In comparing Tables W4A and W4B, note that the recoding of X did not affect either the estimates or the tests of the slope for X or either of the products involving X. The last three rows of the two tables are identical. The only change from recoding X is that now the differences

between the three groups are measured at BMI = 25 instead of the substantively meaningless BMI = 0.

Also note that although the data for the large and small quantity groups are the same as before, the product terms $X'Z_1$ and $X'Z_2$ do not show significant interactions. Even though the slopes for large and small quantity groups differ significantly, that is not tested by the present coding. Instead, the slope of each group is tested against the slope for the no opportunity group, which has a slope near zero between the other two slopes. If we want to analyze the difference between the small quantity and large quantity conditions, we would recode Z_2 such that it is coded 1 for no opportunity and 0 for small quantity or large quantity. The coefficient on Z_1 would then represent the difference between the small quantity and large quantity conditions. This would reveal a significant interaction. An analysis plan for Case 3 is given in Table W5.

Case #4: Continuous x Continuous

Spotlight analysis is typically used when the model includes the product of a typically manipulated, typically dichotomous factor and a typically measured, continuous factor. It may just as easily be used whether each factor is manipulated or measured, and when both factors are continuous. Consider a field study extension of the basic design that does not rely on the use of confederates: quantities taken by the observed consumer (no longer confederates) are now continuous and measured rather than dichotomous and manipulated. BMI's of the observed consumer are still continuous and measured. This change in the design clearly has implications for internal validity since observed consumers are no longer randomly assigned to quantity conditions, but we are focused on the statistical analysis. This model is exactly the same as in our original example of the 2 x Continuous variant of McFerran et al. (2010). The model is:

(W8) $Y = a + bZ + cX + dZX$

Z represents quantity taken, but now it represents a continuous variable, not a dichotomous one. The procedure remains the same and the interpretation of *a*, *c*, and *d* is analogous to before. To interpret the effect of (continuously varying) quantity on quantity taken by the participant, we again shine the spotlight on borderline overweight individuals by setting X' = X – 25 and re-estimating the model Y = *a'* + *b'*Z + *c'*X' + *d'*ZX'. Here, *b'* represents the effect of a one unit change in quantity taken by the observed consumer on quantity taken by the participant, holding constant the observed consumer's BMI at borderline overweight. (There is no analysis template for Case #4 or Case #5 because the same principles from Table W1 apply.)

Case #5: Quadratic

An important but not immediately obvious application of the Continuous x Continuous case is that it can aid in the interpretation when the simple slope in a quadratic model is significant. A quadratic model is essentially a variable interacting with itself. Imagine that we are examining the effect of quantity taken by thin confederates. We might find that assimilation is attenuated at very high levels of quantity taken, and so we are interested in how the effect of the marginal piece of candy taken varies as a function of how many candies are taken by the confederate. Our model would be:

(W9) $Y = a + bZ + cZ^2$.

The simple effect of Z (quantity taken as a continuous variable) is its derivative: $b + 2cZ$. Once again, the coefficient *b* represents the simple effect when Z is equal to 0. If we are interested in the effect of taking one more candy when the confederate has already taken 10 candies, we can calculate Z' = Z – 10 and regress Y on Z' and Z'^2. The coefficient on Z' represents the marginal effect when Z' = 0 or, equivalently, when Z = 10. One can interpret this simple effect as testing the slope of the line tangent to the curve for Z = 10.

BASE CASE 0: SIMPLE EFFECTS IN A 2 x CONTINUOUS DESIGN

$$Y = a + bZ + cX + dZX$$

A. BASELINE ANALYSIS

	Intercept	Manipulation	Measured Variable	Manipulation * Measured
		Z	X	ZX
Coding		0 = Control 1 = Treatment	Raw scale	
Coefficient	*a*	*b*	*c*	*d*
Interpretation	Estimate of Y when Z = 0 and X = 0, i.e., for Control group when X = 0	Simple effect of Treatment vs. Control when X = 0	Simple slope of Measured Variable on Y when Z = 0, i.e., for Control group	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

B. TEST THE SIMPLE EFFECT OF TREATMENT VS. CONTROL AT FOCAL VALUE $X = X_{Focal}$ BY RECODING X SO THAT IT DROPS OUT OF THE EQUATION

		Z	X'	ZX'
Coding		0 = Control 1 = Treatment	$X' = X - X_{Focal}$	
Coefficient	*a'*	*b'*	*c'*	*d'*
Equivalent to	$a + cX_{Focal}$	$b + dX_{Focal}$	*c*	*d*
Interpretation	Estimate of Y when Z = 0 and X'=0, i.e., for Control group when $X = X_{Focal}$	Simple effect of Treatment vs. Control when X' = 0, i.e., when $X = X_{Focal}$	Simple slope of Measured Variable on Y when Z = 0, i.e., for Control group	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

C. TEST THE SIMPLE SLOPE OF X IN TREATMENT GROUP BY RECODING Z SO THAT IT DROPS OUT OF THE EQUATION

		Z"	X	Z"X
Coding		1 = Control 0 = Treatment	Raw scale	
Coefficient	*a"*	*b"*	*c"*	*d"*
Equivalent to	*a + b*	*-b*	*c + d*	*-d*
Interpretation	Estimate of Y when Z" = 0 and X = 0, i.e., for Treatment group when X = 0	Simple effect of Treatment vs. Control when X = 0	Simple slope of Measured Variable on Y when Z" = 0, i.e., for Treatment group	Difference in slope of Measured Variable between Control (Z" = 1) and Treatment (Z" = 0)

D. THE CONSEQUENCES OF USING CONTRAST CODES RATHER THAN DUMMY CODES

		Z’’’	X	Z’’’X
Coding		-1 = Control 1 = Treatment	Raw scale	
Coefficient	*a’’’*	*b’’’*	*c’’’*	*d’’’*
Equivalent to	$a + b/2$	$b/2$	$c + d/2$	$d/2$
Interpretation	Estimate of Y when Z’’’ = 0 and X = 0, i.e., unweighted average of group estimates when X = 0	Half of simple effect of Treatment vs. Control when X = 0	Simple slope of Measured Variable on Y when Z’’’ = 0, i.e., unweighted average of group slopes	Half of difference in slope of Measured Variable between Control (Z’’’ = -1) and Treatment (Z’’’ = 1)

Table W1. Simple effects in a 2 x Continuous design.

CASE 1: SIMPLE EFFECTS IN A 2 (WITHIN) x CONTINUOUS DESIGN

$$Y_2 - Y_1 = a + bX$$

A. BASELINE ANALYSIS

	Intercept	Measured Variable
		X
Coding		Raw scale
Coefficient	*a*	*b*
Interpretation	Estimate of $Y_2 - Y_1$ when X = 0, i.e., Simple effect of Treatment vs. Control when X = 0	Slope of Measured Variable on ($Y_2 - Y_1$), i.e., Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

B. TEST THE SIMPLE EFFECT OF MANIPULATION AT FOCAL VALUE $X = X_{Focal}$ BY RECODING X SO THAT IT DROPS OUT OF THE EQUATION

		X’
Coding		$X' = X - X_{Focal}$
Coefficient	*a’*	*b’*
Equivalent to	$a + bX_{Focal}$	*b*
Interpretation	Estimate of $Y_2 - Y_1$ when X’ = 0, i.e., Simple effect of Treatment vs. Control when $X = X_{Focal}$	Slope of Measured Variable on ($Y_2 - Y_1$), i.e., Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit

Table W2. Simple effects in a 2 (Within) x Continuous design.

CASE 2: SIMPLE EFFECTS IN A 2 x 2 x CONTINUOUS DESIGN

$$Y = a + bZ + cW + dX + eZW + fZX + gWX + hZWX$$

A. BASELINE ANALYSIS

	Intercept	Manipulation 1	Manipulation 2	Measured Variable	Manipulation 1 * Manipulation 2	Manipulation 1 * Measured Variable	Manipulation 2 * Measured Variable	Manipulation 1 * Manipulation 2 * Measured Variable
		Z	W	X	ZW	ZX	WX	ZWX
Coding		0 = Control 1 = Treatment	0 = Group A 1 = Group B	Raw scale				
Coefficient	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Interpretation	Estimate of Y when Z = 0, W = 0, X = 0, i.e., for Control Group A when X = 0	Simple simple effect of Treatment vs. Control when W = 0, X = 0, i.e., simple simple effect of Manipulation 1 for Group A when X = 0	Simple simple effect of Group B vs. Group A when Z = 0, X = 0, i.e., simple simple effect of Manipulation 2 for the Control group when X = 0	Simple simple slope of Measured Variable on Y when Z = 0, W = 0, i.e., for Control Group A	Simple interaction of Manipulation 1 x Manipulation 2 when X = 0	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit for W = 0, i.e., for Group A.	Change in effect of Group B vs. Group A when Measured Variable increases by 1 unit for Z = 0, i.e., for the Control group.	Change in Manipulation 1 x Manipulation 2 interaction when Measured Variable increases by 1 unit

B. TEST THE SIMPLE INTERACTION OF MANIPULATION 1 (Z) x MANIPULATION 2 (W) AT FOCAL VALUE $X = X_{Focal}$ BY RECODING X SO THAT IT DROPS OUT OF THE EQUATION

		Z	W	X'	ZW	ZX'	WX'	ZWX'
Coding		0 = Control 1 = Treatment	0 = Group A 1 = Group B	$X' = X - X_{Focal}$				
Coefficient	*a'*	*b'*	*c'*	*d'*	*e'*	*f'*	*g'*	*h'*
Equivalent to	$a + dX_{Focal}$	$b + fX_{Focal}$	$c + gX_{Focal}$	*d*	$e + hX_{Focal}$	*f*	*g*	*h*
Interpretation	Estimate of Y when Z = 0, W = 0, X' = 0, i.e., for Control Group A when $X = X_{Focal}$	Simple simple effect of Treatment vs. Control when W = 0, X' = 0, i.e., simple simple effect of Manipulation 1 for Group A when $X = X_{Focal}$	Simple simple effect of Group B vs. Group A when Z = 0, X' = 0, i.e., simple simple effect of Manipulation 2 for the Control group when $X = X_{Focal}$	Simple simple slope of Measured Variable on Y when Z = 0, W = 0, i.e., for Control Group A	Simple interaction of Manipulation 1 x Manipulation 2 when X' = 0, i.e., when $X = X_{Focal}$	Change in effect of Treatment vs. Control when Measured Variable increases by 1 unit for W = 0, i.e., for Group A.	Change in effect of Group B vs. Group A when Measured Variable increases by 1 unit for Z = 0, i.e., for the Control group.	Change in Manipulation 1 x Manipulation 2 interaction when Measured Variable increases by 1 unit

C. TEST THE SIMPLE INTERACTION OF MANIPULATION 1 (Z) BY X AT LEVEL W = W_{New0} BY RECODING W SO THAT IT IS 0 FOR THAT LEVEL SO THAT W DROPS OUT OF THE EQUATION

		Z	W”	X	ZW”	ZX	W”X	ZW”X
Coding		0 = Control 1 = Treatment	1 = Group A 0 = Group B	Raw scale				
Coefficient	*a”*	*b”*	*c”*	*d”*	*e”*	*f”*	*g”*	*h”*
Equivalent to	*a + c*	*b + e*	*-c*	*d + g*	*-e*	*f + h*	*-g*	*-h*
Interpretation	Estimate of Y when Z = 0, W” = 0, X = 0, i.e., for Control Group B when X = 0	Simple simple effect of Treatment vs. Control when W” = 0, X = 0, i.e., simple simple effect of Manipulation 1 for Group B when X = 0	Simple simple effect of Group A vs. Group B when Z = 0, X = 0, i.e., simple simple effect of Manipulation 2 for the Control group when X = 0	Simple simple slope of Measured Variable on Y when Z = 0, W” = 0, i.e., for Control Group B	Simple interaction of Manipulation 1 x Manipulation 2 when X = 0	Difference in slope of Measured Variable between Treatment (Z = 1) and Control (Z = 0) when W" = 0, i.e., for Group B.	Difference in slope of Measured Variable between Group A (W" = 1) and Group B (W" = 0) when Z = 0, i.e., for the Control group.	Difference in Group A vs. Group B difference in slope of Measured Variable between Treatment and Control

Note—In conducting an ANOVA of three dichotomous variables, one will typically report main effects, two-way interactions, and the three-way interaction. In the analysis above, we emphasize simple-simple effects and simple interactions. In the 2 x 2 x continuous case, one may easily elicit the analogous overall effect terms, rather than simple effect terms, through judicious recoding such that 0 represents the average rather than one condition or a focal point. These are of particular interest if higher-order interactions are not statistically significantly different from 0.

- To examine the effect of one variable averaged across the sample (analogous to a main effect in ANOVA), mean-center all interacting variables. For example, given equal cell sizes, to examine the effect of X averaged across the sample, code Z such that -1 = Control, 1 = Treatment and code W such that -1 = Group A, 1 = Group B. Note that each coefficient reflects a one-unit change, so the difference between two contrast-coded conditions (2 units) is twice the coefficient (as shown in Table W1D).
- To examine the effect of a two-way interaction averaged across the sample (analogous to a two-way interaction in ANOVA), mean-center the interacting variable. For example, to examine the interaction effect of X and W averaged across the sample, mean-center Z, which in the equal N case, leads to contrast codes -1 = Control, 1 = Treatment.
- To examine the interaction effect of Z and W, mean-center X.
- None of these recoding schemes impacts the implications of the three-way interaction term, though its estimate may nominally change if one scale has been expanded (e.g., 0, 1 to -1, 1) or reversed (e.g., 0, 1 to 1, 0).

Table W3. Simple effects in a 2 x 2 x Continuous design.

A.

Variable	Coefficient	Estimate	Standard Error	*t*	*p*
intercept	a	9.20	3.51	2.62	0.01
small quantity taken	b_1	-5.56	5.06	2.57	0.27
large quantity taken	b_2	13.01	5.05	2.57	0.01
confederate BMI	c	0.02	0.16	0.12	0.90
small quantity taken * confederate BMI	d_1	0.17	0.23	0.75	0.45
large quantity taken * confederate BMI	d_2	-0.40	0.23	-1.70	0.09

B.

Variable	Coefficient	Estimate	Standard Error	*t*	*p*
intercept	a'	9.71	0.74	12.20	0.0001
small quantity taken	b_1'	-1.27	0.97	-1.30	0.19
large quantity taken	b_2'	3.12	1.05	2.98	0.003
confederate BMI – 25	c'	0.02	0.16	0.12	0.90
small quantity taken * (confederate BMI – 25)	d_1'	0.17	0.23	0.75	0.45
large quantity taken * (confederate BMI – 25)	d_2'	-0.40	0.23	-1.70	0.09

Table W4. Regression results of fictitious illustrative data for the 3 x Continuous case, where the confederate takes either 2 candies ($Z_1 = 1$, $Z_2 = 0$), 30 Candies ($Z_1 = 0$, $Z_2 = 1$), or the confederate does not have an opportunity to take any candies ($Z_1 = 0$, $Z_2 = 0$). Panel A: X is the confederate's BMI. Panel B: X' is the confederate's BMI – 25.

CASE 3: SIMPLE EFFECTS IN A 3 x CONTINUOUS DESIGN

$$Y = a + b_1Z_1 + b_2Z_2 + cX + d_1Z_1X + d_2Z_2X$$

A. BASELINE ANALYSIS

	Intercept	Manipulation	Manipulation	Measured Variable	Manipulation * Measured Variable	Manipulation * Measured Variable
		Z_1	Z_2	X	Z_1X	Z_2X
Coding		0 = Control 1 = Treatment 1 0 = Treatment 2	0 = Control 0 = Treatment 1 1 = Treatment 2	Raw scale		
Coefficient	a	b_1	b_2	c	d_1	d_2
Interpretation	Estimate of Y when $Z_1 = 0$, $Z_2 = 0$, and X = 0, i.e., for Control group when X = 0	Simple effect of Treatment 1 vs. Control when X = 0	Simple effect of Treatment 2 vs. Control when X = 0	Simple slope of Measured Variable on Y when $Z_1 = 0$ and $Z_2 = 0$, i.e., for Control group	Change in effect of Treatment 1 vs. Control when Measured Variable increases by 1 unit	Change in effect of Treatment 2 vs. Control when Measured Variable increases by 1 unit

B. TEST THE SIMPLE EFFECT OF TREATMENT 1 VS. CONTROL AT FOCAL VALUE $X = X_{Focal}$ BY RECODING X SO THAT IT DROPS OUT OF THE EQUATION

		Z_1	Z_2	X’	Z_1X'	Z_2X'
Coding		0 = Control 1 = Treatment 1 0 = Treatment 2	0 = Control 0 = Treatment 1 1 = Treatment 2	$X' = X - X_{Focal}$		
Coefficient	a'	b_1'	b_2'	c'	d_1'	d_2'
Equivalent to	$a + cX_{Focal}$	$b_1 + d_1X_{Focal}$	$b_2 + d_2X_{Focal}$	c	d_1	d_2
Interpretation	Estimate of Y when $Z_1 = 0$, $Z_2 = 0$, and X’ = 0, i.e., for Control group when $X = X_{Focal}$	Simple effect of Treatment 1 vs. Control when X’ = 0, i.e., when $X = X_{Focal}$	Simple effect of Treatment 2 vs. Control when X’ = 0, i.e., when $X = X_{Focal}$	Simple slope of Measured Variable on Y when $Z_1 = 0$ and $Z_2 = 0$, i.e., for Control group	Change in effect of Treatment 1 vs. Control when Measured Variable increases by 1 unit	Change in effect of Treatment 2 vs. Control when Measured Variable increases by 1 unit

C. TEST THE SIMPLE SLOPE OF X IN TREATMENT 1 BY RECODING Z_1 SO THAT IT DROPS OUT OF THE EQUATION

		Z_1"	Z_2	X	Z_1"X	Z_2X
Coding		1 = Control 0 = Treatment 1 0 = Treatment 2	0 = Control 0 = Treatment 1 1 = Treatment 2	Raw scale		
Coefficient	a"	b_1"	b_2"	c"	d_1"	d_2"
Equivalent to	$a + b_1$	$-b_1$	$b_2 - b_1$	$c + d_1$	$-d_1$	$d_2 - d_1$
Interpretation	Estimate of Y when $Z_1 = 0$, $Z_2 = 0$, and X = 0, i.e., for Treatment 1 group when X = 0	Simple effect of Control vs. Treatment 1 when X = 0	Simple effect of Treatment 2 vs. Treatment 1 when X = 0	Simple slope of Measured Variable on Y when $Z_1 = 0$ and $Z_2 = 0$, i.e., for Treatment 1 group	Difference in slope of Measured Variable between Control (Z_1" = 1) and Treatment 1 (Z_1" = 0)	Difference in slope of Measured Variable between Treatment 2 ($Z_2 = 1$) and Treatment 1 ($Z_2 = 0$)

Table W5. Simple effects in a 3 x Continuous design.

Web Appendix B

Why Not Test at Plus and Minus One Standard Deviation?

Researchers most commonly will first mean-center their continuous measure, and then transform it to spotlight the effect of the manipulation at plus and minus one standard deviation from the mean; these values were suggested by Cohen and Cohen (1983) for cases where there are no substantively meaningful values. This approach is not wrong, but we would argue that it is suboptimal and we cannot generate a case where it would be the preferred approach. Using the BMI example, depending on the sample, one standard deviation above the mean might be "normal" weight or it might be clinically obese. Moreover, it is hard to argue that we should be more interested in the effect of Z at exactly one standard deviation above the mean of X in this particular sample than in values slightly higher or slightly lower.

There are three main problems of testing at plus and minus one standard deviation. First, if the distribution of the moderator X is skewed, one of those values can be outside the range of the data. Second, if the moderator X is on a coarse scale, it may be impossible to have a value of X exactly equal to plus or minus one standard deviation. Third, if two researchers replicate the same study with samples of very different mean levels of the moderator, it can appear that they fail to replicate each other even when they find exactly the same regression equation in raw score units. This problem is exacerbated by the tendency for authors using the plus and minus one standard deviation approach to fail to report the mean and standard deviation of X.

As an example of these issues and how to solve them, Fernbach et al. (2013) exposed respondents to product concepts that gave high, medium, or low levels of causal detail for why and how a new product delivered a claimed benefit and asked how well respondents understood the concept. Fernbach et al. showed that people who score high on Frederick's (2005) Cognitive

Reflection Test (CRT) believed that they understood the concept better when more causal detail was given, but people who scored low believed they understood the concept better when low causal detail was given. The CRT scale is a 3-item quiz resulting in a score of 0, 1, 2, or 3 questions right. Fernbach et al. chose to report the test of the simple effect of the manipulation of causal detail at real possible scores of 0 and 3 rather than to report tests at fractional values of CRT at plus and minus one standard deviation from the mean. Fractional scores are impossible for any individual participant to receive.

Moreover, in this particular example the distribution of CRT is skewed, so it would be easy to have a value one standard deviation below the mean that is actually below zero in a particular sample. It would have been meaningless to test such a simple effect that is outside the range of the data. For example, Frederick (2005) reports distributions of CRT scores across several different institutions. Had Fernbach et al. (2013) tested their hypothesis one standard deviation above and below the mean with Frederick's MIT sample ($M = 2.18$, $SD = 0.94$), one standard deviation above the mean would have been an impossibly high value. If they tested their hypothesis one standard deviation above and below the mean with Frederick's University of Toledo sample ($M = 0.57$, $SD = 0.87$), one standard deviation below the mean would have been an impossibly low value. Unfortunately, the modal reporting strategy is to omit reporting the distribution of the individual difference measure in the sample. Had Fernbach conducted one study at MIT and one at Toledo and examined simple effects at one standard deviation above and below the means without reporting the sample distribution, the simple effects would have appeared to be inconsistent because the distributions of CRT differed so much. Determining whether results replicated across samples without knowing their distributions would be impossible. This is not a problem for interpreting a study in isolation, but for comparing studies.