Experimental Personality Designs: Analyzing Categorical by Continuous Variable Interactions

Stephen G. West, Leona S. Aiken, and Jennifer L. Krull
Arizona State University

ABSTRACT Theories hypothesizing interactions between a categorical and one or more continuous variables are common in personality research. Traditionally, such hypotheses have been tested using nonoptimal adaptations of analysis of variance (ANOVA). This article describes an alternative multiple regression-based approach that has greater power and protects against spurious conclusions concerning the impact of individual predictors on the outcome in the presence of interactions. We discuss the structuring of the regression equation, the selection of a coding system for the categorical variable, and the importance of centering the continuous variable. We present in detail the interpretation of the effects of both individual predictors and their interactions as a function of the coding system selected for the categorical variable. We illustrate two- and three-dimensional graphical displays of the results and present methods for conducting post hoc tests following a significant interaction. The application of multiple regression techniques is illustrated through the analysis of two data sets. We show how multiple regression can produce all of the information provided by traditional but less optimal ANOVA procedures.

Stephen G. West was partially supported by NIMH Grant P50-MH39246 and a sabbatical leave from Arizona State University to study at UCLA during the writing of this article. We thank Michael Kernis for making the data reported in Example 2 available to us. We thank the participants in the 1994 Nags Head Conference on Personality and Social Behavior for their suggestions. We also thank Khanh Bui, action editor William Chaplin, Eileen Donahue, David Funder, William Graziano, Gary McClelland, Julie Norem, and several anonymous reviewers for their helpful comments on an earlier version of the manuscript. Graphical help and suggestions from Patrick Curran, William Mason, Gary McClelland, and Robert Weiss are also acknowledged. Correspondence concerning this article should be addressed to Stephen G. West, Department of Psychology, Arizona State University, Box 87-1104, Tempe, AZ 85287-1104.

Journal of Personality 64:1, March 1996.

Personality researchers are frequently faced with the analysis of designs involving both categorical and continuous variables. Most commonly, these issues arise in the context of the "experimental personality" (aptitude-treatment) design in which each subject is initially measured on one or more individual difference measures. Subjects are then randomly assigned to treatments within an experiment and their responses on the outcome variable are measured (see Atkinson & Feather, 1966; Kernis, Cornell, Sun, Berry, & Harlow, 1993; Rotter & Mulry, 1965, for examples). Interactionist theories predict that situational factors modify the relation between personality traits and behavior (Krahé, 1992; Magnusson & Endler, 1977; Snyder & Ickes, 1985). Similar analytic issues also arise in designs in which natural categories like gender are expected to interact with continuous variables in predicting an outcome.

Traditionally, the data from such designs were analyzed using the familiar framework of analysis of variance (ANOVA). To illustrate, consider the design and analysis used by Rotter and Mulry (1965). In their design, each subject's locus of control was measured using Rotter's (1966) I-E scale, the description of the nature of the task (chance- vs. skill-based) was manipulated, and the subject's decision time on a pattern-matching task was recorded as the outcome variable. The data from the continuous, individual difference variable (locus of control) were divided at the median into a "high" (internal) and "low" (external) group. The data on the outcome (dependent) variable of decision time were then analyzed using a 2 × 2 (Task Description × Locus of Control) ANOVA, yielding tests of the main effect of locus of control, the main effect of task description, and the Locus of Control × Task Description interaction, each with 1 degree of freedom in the numerator. Well-developed prescriptions within the ANOVA approach allowed for the interpretation of the effects, graphical presentation of the results, and post hoc probing of significant interactions through tests of simple effects (Winer, 1971; Winer, Brown, & Michels, 1991). We will term this approach "ANOVA with cutpoints."

Unfortunately, a number of problems have become evident in recent years with the application of the traditional ANOVA with cutpoints approach. First, Cohen (1983) illustrated how artificially dichotomizing a continuous variable greatly reduces the power of statistical tests. In the experimental personality design, this problem affects both the tests of the individual difference variable(s) and the interaction(s) of the individual difference and manipulated variables. This problem is particularly important given the generally low power of tests of interactions involving continuous variables (Aiken & West, 1991; Chaplin, 1991; Cronbach & Snow, 1977; McClelland & Judd, 1993; Stone-Romero & Anderson, 1994). Second, Maxwell and Delany (1993) and Pitts and West (1995) have shown that an even more serious problem emerges in designs in which there are two or more correlated individual difference variables: Statistically significant but completely spurious effects of the individual difference variables may be detected, even when the individual difference variables, in fact, have no relation to the outcome variable. Other problems also exist, notably the very limited ability of the ANOVA with cutpoints approach to detect model misspecification (e.g., curvilinear effects of continuous variables; see Aiken & West, 1991, chaps. 5 and 9), but will not be the focus of the presentation here.

Because of these problems, the traditional ANOVA with cutpoints approach is being increasingly superseded by a multiple regression-based approach, often termed *moderated multiple regression* (Aiken & West, 1991; Jaccard, Turrisi, & Wan, 1990; Judd & McClelland, 1989; Saunders, 1956). As originally outlined by Cohen (1968; see also Cohen & Cohen, 1983), any combination of continuous and categorical predictor variables can be analyzed within a multiple regression framework. Categorical variables are represented through one or more code variables that assign a unique value to each group (e.g., dummy codes such as male $= 0$, female $= 1$, or unweighted effects codes such as male $= -1$, female $= +1$). Interactions are represented as the products of individual predictors (e.g., the product XW of two predictors X and W). Curvilinear relations are represented through higher order functions of predictors in the regression equation (e.g., the square of a predictor, X^2, for a U-shaped relationship).

The purpose of this article is to inform readers about the use of the multiple regression approach, particularly as it applies to the analysis of experimental personality designs. We extend the work of Aiken and West (1991) by focusing our presentation on designs involving two individual difference variables and one manipulated situational variable. Such designs are important in both classic and contemporary work in personality. For example, Atkinson and Feather's (1966) theory of achievement motivation predicts a complex interaction among need for achievement (continuous), fear of failure (continuous), and task success versus failure (experimentally manipulated) in determining future performance. Recent work on self-esteem (e.g., Campbell, 1993; Kernis et al., 1993) predicts interactions among a person's chronic level of self-esteem, the stability of their level of self-esteem, and situational factors in determining the subject's responses. The methods presented in this article can be generalized both upward to still more complex designs involving additional manipulated or measured independent variables, as well as downward to the simpler, classic experimental personality design involving only one manipulated and measured variable (see Aiken & West, 1991, chap. 7). Although our focus will be on the experimental personality design, we will also consider designs in which the categorical predictor represents naturally occurring (e.g., gender) rather than manipulated groups. Space considerations limit our presentation to between-subjects designs in which each subject is exposed to a single treatment condition and is measured only once on the outcome variable.

In this article we present a step by step approach to the analysis of experimental personality designs.

1. We begin by exploring how to structure regression equations containing two continuous predictors that interact with a two-group and then a three-group categorical variable.

2. We show several potential systems of code variables for representing the categorical independent variable in regression equations.

3. We discuss the importance of centering (i.e., putting in deviation score form) continuous variables when interactions occur in regression equations.

4. We explain how to conduct omnibus tests of the categorical variable itself and of interaction effects involving the categorical variable.

We then present two example analyses.

5. Example 1 includes a categorical variable with three levels and illustrates the general interpretation of regression coefficients associated with each of the systems of creating code variables. The discussion of Example 1 also considers issues in selecting among the coding systems.

6. Example 2, involving real data, illustrates simplifications that occur in the regression model when there are only two levels of the categorical variable. Example 2 also introduces methods for presenting the results using two- and three-dimensional graphical displays and for post hoc testing of significant interactions.

Our goal in this article is to provide a comprehensive set of prescriptions for the analysis, interpretation, graphical display, and post hoc probing of significant interactions within the multiple regression framework. These prescriptions provide parallels to all of the information available in ANOVA with categorical variables.

Structuring the Regression Equation

Complete factorial ANOVA models, with which psychologists are most familiar, are structured so that all main effects and two-way interactions involved in a three-way interaction are included in the model. With factors A, B, and C, we have the main effect terms A, B, and C; the two-way interactions AB, AC, BC; and the three-way interaction ABC. Regression models involving interactions must be structured in the same manner as are complete factorial ANOVA models. Each variable included in the highest order interaction must serve as a separate predictor, and all possible combinations of the individual predictors that are contained in the highest order interaction must also serve as predictors in the regression equation. Such regression equations are described as being hierarchically well formulated (Peixoto, 1987).

First, consider a regression equation with continuous predictors X and W, and a two-group categorical variable C (e.g., gender, represented using unweighted effect codes of -1 = male and $+1$ = female):

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C + b_4XW + b_5XC + b_6WC + b_7XWC \qquad (1)$$

The correspondence between Equation 1 and the familiar three-factor ANOVA model is apparent. The *highest order interaction* is the three-way XWC interaction. The *lower order interactions* include three two-way interactions: between the continuous variables X and W, between continuous variable X and the categorical (code) variable C, and between continuous variable W and categorical variable C. The *first-order effects* of each predictor, X, W, and C in this illustration, are analogous to main effects in ANOVA. But recall that most ANOVA texts contain admonitions against interpreting main effects in the presence of significant interactions. For this reason, we will refer to the effects of individual predictors in regression equations containing interactions as first-order effects (rather than as main effects) throughout this article. Later, we will develop interpretations of first-order effects as "average" or "conditional" effects, interpretations that remain useful even in the presence of higher order interactions.

Second, consider a case in which the categorical variable has $G = 3$ levels. When the categorical variable has $G > 2$ levels, then more than one code variable must be built into the regression equation to fully represent the categorical variable. With G groups in all, we need $(G-1)$ code variables. For our second case, we need two code variables, which we will name C_1 and C_2, to represent the three levels of the categorical variable. These two code variables, C_1 and C_2, taken together represent the first-order effect of the categorical variable. To form the interaction of the categorical variable with a continuous variable X, two cross-product terms are required: XC_1 and XC_2. These two terms taken together represent the two-way interaction of X by the categorical variable. Otherwise stated, C_1 and C_2 form a set that represents the three groups; both code variables must be included in the equation to represent the first-order effect of the group variable and its interaction with other variables. A full regression equation containing the interaction of continuous variables X and W with a three-level categorical variable would be structured as in Equation 2 below:

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_6XC_1 + b_7XC_2 + b_8WC_1 + b_9WC_2 + b_{10}XWC_1 + b_{11}XWC_2 \quad (2)$$

We will refer to Equation 2 frequently throughout this article.

The Categorical Independent Variable: Choice of Coding System

Once the regression equation has been structured, the next step in the analysis is to choose a coding system with which to represent the categorical variable. We require a set of $(G-1)$ code variables to represent the differences among the G groups in a regression analysis. In practice, personality researchers will typically mount only $G = 2$ or $G = 3$ different manipulations (or use a similarly small number of natural categories) and so will require only one or two code variables, respectively. Each of the coding systems leads to a different interpretation that may or may not be optimal, depending on the regression model being considered and the questions being posed by the researcher (Aiken & West, 1991; Cohen & Cohen, 1983; Pedhazur, 1982; Serlin & Levin, 1985; Suits, 1984).

Coding systems consist of numerical values assigned to members of different levels of the categorical variable (e.g., female = 1; male = 0). Two fundamental questions should be addressed for each coding system:

(*a*) What is the meaning of a value of 0 for each code variable? As we will clarify later in this article, when a regression equation contains interactions, each first-order effect and lower order interaction represents the regression of the dependent variable on the predictor or interaction among predictors at the values of 0 for all remaining predictors in the equation. Thus, the meaning of 0 for each code variable will have important implications for the interpretation of the regression coefficients.

(*b*) What is the meaning of a 1-unit change in the code variable? Again, as we will clarify below, the unstandardized regression coefficients for the code variables provide direct estimates of the difference between a mean or slope associated with a specific treatment group and the value represented by 0 (typically a grand mean or the mean of a comparison group) for that code variable.

Throughout this article, we use *unstandardized* regression coefficients because they have a straightforward interpretation within the coding systems we will describe for categorical variables. In regression equations without interactions, standardized coefficients are interpreted in terms of z scores, which rarely have a useful interpretation for categorical variables. In standardized equations including interactions, the coefficients for the interaction terms are *not* properly standardized and are therefore *not* interpretable (Aiken & West, 1991, pp. 40–47; Friedrich, 1982).

Coding systems for categorical variables

We consider four coding systems for categorical variables: dummy codes, unweighted effect codes, weighted effect codes, and contrast codes.

Dummy codes. The top section (A) of Table 1 illustrates three versions of the dummy variable coding system for the three-group case. In this familiar system, a comparison group (Group G) is designated and is assigned a value of 0 for each code variable. The choice of the comparison group is statistically arbitrary; however, there are three practical considerations that should guide this choice: (*a*) The comparison group should in some way serve as a base group in the design (e.g., a control group; a standard treatment; the group expected to score lowest or highest on the dependent variable); (*b*) the comparison group should be well defined (i.e., not a wastebasket category such as "other" for religion); and (*c*) the comparison group ideally should *not* have a very

Table 1
Illustration of Coding Systems for Categorical Variables

	Group 1 as Base		Group 2 as Base		Group 3 as Base	
(A) Dummy codes	C_1	C_2	C_1	C_2	C_1	C_2
Group 1	0	0	1	0	1	0
Group 2	1	0	0	0	0	1
Group 3	0	1	0	1	0	0
(B) Unweighted effects codes	C_1	C_2				
Group 1	1	0				
Group 2	0	1				
Group 3	−1	−1				
	General form		Illustration			
(C) Weighted effects codes[a]	C_1	C_2	C_1	C_2		
Group 1	1	0	1	0		
Group 2	0	1	0	1		
Group 3	$-n_1/n_3$	$-n_2/n_3$	−60/120	−220/120		
(D) Contrast codes[b]	C_1	C_2				
Group 1	+1/3	−1/2				
Group 2	+1/3	+1/2				
Group 3	−2/3	0				

a. n_1, n_2, and n_3 are the sample sizes in the corresponding groups. To illustrate, the group sizes in Example 1 are 60, 220, and 120, respectively.

b. The two contrast codes depicted compare: the unweighted mean of Groups 1 and 2 with the mean of Group 3; the mean of Group 1 with the mean of Group 2.

small sample size relative to the other groups (see Hardy, 1993). For the presentation below, we chose Group 3 as our comparison group and use the set of dummy codes presented in the third column of Table 1(A).

Each other group in turn is given a value of 1 on the code variable that will contrast it with the comparison group and a value of 0 otherwise. As illustrated in Table 1(A) using Group 3 as the base, C_1 contrasts Group 1 with Comparison Group 3 and C_2 contrasts Group 2 with Comparison Group 3. All $(G - 1)$ code variables must be included in the regression equation to represent the overall treatment effect. If some of a set of code variables representing the treatment are not included in the regression equation, the interpretation of the regression coefficients changes, often in a dramatic manner (see Serlin & Levin, 1985). In the regression equation, each code variable contributes 1 degree of freedom (df) as a predictor. The set of $(G - 1)$ code variables have $(G - 1)$ df in all, completely equivalent to the $(G - 1)$ degrees of freedom for the main effect of a categorical variable with G levels in ANOVA.

Consider Equation 3 below, a simple regression equation comparing the means of three treatment groups,

$$\hat{Y} = b_0 + b_1C_1 + b_2C_2 \tag{3}$$

In this equation, $\hat{Y}$ is the predicted value of the outcome variable, b_0 is the intercept, b_1 is the coefficient for the first dummy code (C_1), and b_2 is the regression coefficient for the second dummy code (C_2). Each of these regression coefficients is unstandardized. If we substitute the values of the dummy codes corresponding to each group into Equation (3), we find:

Group 1: $\hat{Y} = b_0 + b_1(1) + b_2(0) = b_0 + b_1 = M_1$
Group 2: $\hat{Y} = b_0 + b_1(0) + b_2(1) = b_0 + b_2 = M_2$
Group 3: $\hat{Y} = b_0 + b_1(0) + b_2(0) = b_0 = M_3$
(Group 3 is the comparison group)

Thus, in Equation 3 when $C_1 = 0$ and $C_2 = 0$, $\hat{Y}$, the predicted value of the outcome variable, equals b_0, the regression intercept, which also equals M_3, the mean of the comparison group. The same $\hat{Y}$ value is predicted for all subjects in the comparison group. Correspondingly, a 1-unit change on C_1 (i.e., a change of the value of C_1 from 0 for the comparison group to a value of 1 for Group 1) represents the difference in the value of the Group 1 mean and the comparison group mean on the outcome variable. A 1-unit change on C_2 is associated with the difference in the means of Group 2 and the comparison group on the outcome variable.

Other dummy variable-like coding systems can be developed that assign numbers other than 0 to the comparison group and numbers other than 1 to represent group membership (e.g., $C_1 = 1$ for male; $C_2 = 2$ for female). Such coding systems produce similar results to standard dummy coding in regression models without interactions. However, some can produce results that are far more difficult to interpret with regression models involving interactions because the meaning of the regression coefficients changes. Such coding systems are *not* recommended.

Unweighted effects codes. In unweighted effects codes, a base group is arbitrarily designated and is assigned a value of -1 for each code. Each of the other treatment groups is assigned a value of $+1$ for one code and a value of 0 for all other codes. This coding system is illustrated in the second section of Table 1 for our three-group example, arbitrarily designating Group 3 as the base group. Substituting the values of the unweighted effect codes corresponding to each group into the regression equation, we find:

$$\begin{aligned}
&\text{Group 1: } \hat{Y} = b_0 + b_1(1) + b_2(0) = b_0 + b_1 = M_1\\
&\text{Group 2: } \hat{Y} = b_0 + b_1(0) + b_2(1) = b_0 + b_2 = M_2\\
&\text{Group 3: } \hat{Y} = b_0 + b_1(-1) + b_2(-1) = b_0 - b_1 - b_2 = M_3
\end{aligned}$$

In this coding system, b_0 represents the *unweighted* grand mean of all of the groups $[M_u = (M_1 + M_2 + M_3)/3]$. In fact, all regression coefficients for individual group codes represent discrepancies from the unweighted mean when unweighted effects codes are used. The group coded $[-1, -1]$ is *not* a comparison group against which other groups are compared in this coding scheme. b_1 represents the change in $\hat{Y}$ associated with a 1-unit change on C_1, which equals the difference between the mean of Group 1 and the unweighted grand mean. b_2 represents the change in $\hat{Y}$ associated with a 1-unit change on C_2, which equals the difference between the mean of Group 2 and the unweighted grand mean. The difference between the Group 3 mean and the unweighted grand mean is computed from b_1 and b_2 as $-(b_1 + b_2)$.

Weighted effects codes. Weighted effects codes (Darlington, 1990; Winer et al., 1991) follow the same logic as unweighted effects codes except that the size of each treatment group is taken into consideration. To illustrate, consider a study in which there are three groups that have sample sizes of $n_1 = 60$, $n_2 = 220$, and $n_3 = 120$. Once again, we will arbitrarily designate Group 3 as the base group. The values of C_1 and C_2 follow the same pattern as we observed for unweighted effect codes. However, the values of the code are adjusted (weighted) for Group 3 to reflect the different sample sizes of each of the groups. The left column of the weighted effects codes section (C) of Table 1 presents the general form of weighted effects codes for the three-group case; the right column of Table 1(C) presents the specific values of the codes for the first illustrative study (Example 1) to be described below. The difference in sample sizes is represented in the code of the group that receives the negative value. For each code, this value is *minus* the ratio of the size of the group coded 1 and the size of the group that would have been coded -1 using unweighted effects codes. Note that when sample sizes are equal across groups, $n_1 = n_2 = n_3$, the codes for Group 3 simplify to $[-1, -1]$. Under these conditions, weighted effects codes are identical to unweighted effects codes and produce *identical* results.

Using weighted effects codes, b_0 represents the weighted grand mean (M_w) of the group means, $b_0 = M_w = [(n_1M_1 + n_2M_2 + n_3M_3)/(n_1 + n_2 + n_3)]$. Each regression coefficient in this system represents the difference between the mean of a specified group and the weighted grand mean. b_1 represents the change in $\hat{Y}$ for a 1-unit change on C_1, which is the difference between the Group 1 mean and the weighted grand mean. b_2 represents the change in $\hat{Y}$ for a 1-unit change on C_2, which is the difference between the Group 2 mean and the weighted grand mean. The value $[(-n_1/n_3)b_1 + (-n_2/n_3)b_2]$ is the difference between the Group 3 mean and the weighted grand mean.

Contrast codes. Contrast codes (Judd & McClelland, 1989; Rosenthal & Rosnow, 1985) are the familiar a priori comparisons discussed in traditional ANOVA texts (e.g., Kirk, 1995; Winer et al., 1991). They are used if the researcher has specific, a priori hypotheses that involve linear combinations of two (or more) treatment group means or slopes. As one example, imagine a researcher has a control group (Group 3) and two treatment groups (Group 1, Group 2). The researcher predicts that (a) the mean of the two treatment groups combined will differ from the mean of the control group and (b) the two treatment groups will not themselves differ. The coding system represented in the bottom section

(D) of Table 1 captures these two contrasts.[1] b_0 is the unweighted grand mean of the three groups. b_1 represents the value of a 1-unit change on C_1, which represents the difference between the Group 3 (control) mean and the unweighted mean of Groups 1 and 2 (treatment). The test of b_1 represents the test of hypothesis (a). b_2 represents the value of a 1-unit change on C_2, which represents the difference between the mean of Group 1 and the mean of Group 2. The test of b_2 represents the test of hypothesis (b).

The specific contrasts selected for comparison depend on the researcher's a priori hypotheses. Three rules maximize the interpretability of the contrasts. First, the sum of the weights for each code variable must equal 0. For example, for the first code variable $[(+1/3) + (+1/3) + (-2/3)] = 0$. Second, the difference between the value of the positive weights and negative weights should equal 1 (e.g., $[(+1/3) - (-2/3) = 1]$. This rule ensures that the regression coefficient corresponding to the contrast will directly provide the value of the difference between the unweighted means of the groups involved in the contrast (rather than the value of interest multiplied by a constant). Third, to achieve orthogonal contrasts that account for nonoverlapping variance in the dependent variable when the sample sizes are equal in each group, the sum of the products of each pair of codes should be 0. For example, in Table 1(D), the product of the two codes is $(+1/3)(-1/2) = (-1/6)$ for Group 1, $(+1/3)(+1/2) = (+1/6)$ for Group 2, and $(-2/3)(0) = 0$ for Group 3. The sum of these product terms $(-1/6) + (+1/6) + (0) = 0$. These three rules lead to directly interpretable contrasts whether or not the sample size is equal across groups. However, the contrasts will be orthogonal only when the group sizes are equal.

Contrast analysis offers a useful approach when the researcher has a set of strong a priori predictions. Rosenthal and Rosnow (1985) discuss the philosophy of this general approach and describe its use in the ANOVA context. Judd and McClelland (1989) outline the use of contrast codes in multiple regression analysis, particularly as applied to

1. The contrast codes described here are unweighted in that differences in group sizes are not taken into account. Weighted contrast codes have also been defined (see, e.g., Kirk, 1995, pp. 761–764; Serlin & Levin, 1985). The criteria for choosing unweighted versus weighted contrast codes parallel those for choosing unweighted versus weighted effect codes discussed later in the article. In general, unweighted contrast codes will typically be more useful for experimental personality designs in which the group variable is manipulated.

designs involving experimental manipulations. Serlin and Levin (1985) present a very general matrix-based method for deriving codes that provide weighted or unweighted tests for any set of contrasts of interest.

Comments on coding systems

Each of the coding systems represents a different way of partitioning the information from the G groups. The value of the test statistic of the 1 degree of freedom tests associated with the corresponding terms in each coding system will generally differ when $G > 2$. The results from one coding system can be converted into those of another. For example, in our illustration of dummy codes, the test of b_1 in Equation 3 represents the difference between the means of Groups 1 and 3 ($M_1 - M_3$). For unweighted effects codes, $(2b_1 + b_2)$ represents the difference between the means of Groups 1 and 3. However, it is far simpler to choose the coding system that provides direct answers to the researcher's specific question than to choose one that requires additional, perhaps complex calculations to answer the question.[2]

Continuous Independent Variables: Centering

In regression equations containing interactions, each regression coefficient represents the regression of the dependent variable on the specific variable at the value of 0 on all other variables. Thus, the meaning of the value 0 on each continuous variable must be considered.

Aiken and West (1991, chap. 3) present a detailed discussion of the advantages of centering continuous variables in regression equations. Centering simply means converting each continuous variable to deviation score form, making the mean of the variable 0 while preserving the units of the scale. Centered X is calculated as:

$$X = X_{\text{raw}} - \text{Mean}(X)$$

Centering continuous variables has several advantages that are of particular importance in the analysis of regression models involving interactions.

First, psychological scales rarely have meaningful 0 points. Even on

2. Aiken and West (1991, pp. 24–26) present a general method that can be applied to test any linear combination of coefficients in a regression equation.

well-developed individual difference measures like IQ, a score of 0 does not have a clear meaning. Regression coefficients in complex models involving interactions are *conditional effects*. Conditional effects refer to effects that hold *only* at specific values of other predictors in the equation. First-order effects (e.g., the regression coefficient for continuous variable X) are interpreted when all other continuous variables and codes for categorical variables have a value of 0. Lower order interactions (e.g., the regression coefficient for the interaction between continuous X and continuous W in a model containing three-way interactions) are interpreted at a value of 0 for the third variable. Centering the continuous variables ensures that the interpretation of effects will occur at a meaningful value of the continuous variable (i.e., the mean, which has a value of 0 with centered variables).

Second, centering the continuous variables yields the regression model that is most analogous to the familiar ANOVA model. In ANOVA, we interpret a main effect as the *constant* effect of one factor that holds across all levels of other factors. Consider an equal *n*s two-factor ANOVA in which the A main effect and the AB interaction are both significant. The significance of the AB interaction indicates that the amount of difference between the levels of A depends upon the particular level of factor B at which the difference among levels of A is considered. The A main effect no longer represents a constant effect; rather it represents the *average* amount of discrepancy among the levels of factor A taken across all levels of B. Equivalently, the A main effect represents the amount of discrepancy among the levels of factor A *at the mean of factor* B. When the AB interaction (or any other interaction involving A) is significant, then the A main effect is *conditional* upon the value of the other factors at which it is interpreted; it is no longer the constant effect of factor A that holds regardless of the level of factor B.

In multiple regression analysis precisely the same change in interpretation of first-order effects occurs when the first-order effect is included in an interaction. In an equation containing a group variable G and a continuous variable X, if the continuous variable and the group variable interact, the amount of difference among the groups depends upon the particular value of the continuous variable. When the continuous variable X has been centered, the interpretation of the first-order effect of the group is as the *average* effect of the group variable across all values of the continuous variable. Alternatively, with centered X, the effect of the group variable G can be interpreted as the effect of the group variable at the mean of X, or at the value of 0 on the X variable (since X is centered, its mean is 0). In sum, the interpretations of the effects of first-order variables that also are contained in interactions are identical in ANOVA and multiple regression if the predictors have been centered.[3]

If the continuous variable has not been centered, then the interpretation of the first-order effect of the group variable is different. The first-order effect still represents the amount of difference among the groups at the value of 0 of the continuous variable X; however, 0 is no longer at the mean of X. In fact, 0 may not even exist on the scale, rendering the interpretation of the group effect psychologically meaningless (e.g., if X is a continuous 7-point Likert scale with a range of 1 to 7). This difference in the meaning of the value of 0 at which regression coefficients are interpreted yields disconcerting changes in the magnitude of regression coefficients as predictor variables are rescaled. The regression coefficient for a first-order term or lower order interaction may change from highly negative to highly positive, from significant to nonsignificant. The one exception is that the regression coefficient for the highest order term(s) in the equation remains *constant* across any linear rescaling of the continuous predictors. For example, in Equation 1 the term b_7XWC remains constant and in Equation 2 the terms $b_{10}XWC_1$ and $b_{11}XWC_2$ remain constant.

To illustrate these points, consider an evaluation of a school lunch program relative to a no treatment control that shows an interaction with family income: The poorest children will likely show larger gains in health from the program relative to the better-off children, who show more modest gains. If income has been centered, then the first-order effect of the lunch program can be interpreted as the average effect of the lunch program across all children in this sample. It can also be interpreted as the amount of benefit a child at the mean level of family income in this sample of poor families could expect from the program, often a useful value. In contrast, if income has *not* been centered, then the first-order effect of the lunch program predicts how much benefit children from families with $0 income could expect from the program. Assuming that all sources, including welfare, have been included in the computation of income, this latter value is unlikely to be useful.

Third, many users of multiple regression with interactions have observed a disconcerting result, namely that correlations between first-

3. This statement does not hold for dummy codes since they are evaluated relative to the comparison group rather than to the mean of the groups.

order predictors (e.g., X and Z) and interactions containing those predictors (e.g., XZ) can change dramatically depending on how X and Z are scaled. If X and Z are centered, then the correlation of X and XZ or Z and XZ will typically be low. In contrast, these correlations will often be high if X and Z are uncentered. Marquardt (1980) has distinguished between two sources of multicollinearity in regression equations with interactions. Essential ill-conditioning results from true relationships between variables in the population (e.g., between intelligence and authoritarianism). Nonessential ill-conditioning results from relations between the means of the variables. Centering reduces multicollinearity because it eliminates nonessential (but not essential) ill-conditioning. To illustrate, consider the variable X with five scores: 1, 2, 3, 4, 5. Squaring these scores yields X^2: 1, 4, 9, 16, 25. In a regression equation containing both X and X^2 terms, these two terms will often be highly correlated, $r = .98$ in the present illustration. However, if X and X^2 are centered, the correlation between these terms is expected to be dramatically reduced. In our illustration, mean$(X) = 3$ so that the centered X scores are -2, -1, 0, $+1$, $+2$ and centered X^2 is 4, 1, 0, 1, 4. Recalculation of the correlation between centered X and X^2 yields $r = .00$. This advantage is particularly important in regression equations containing several lower order interactions of interest.[4]

Centering does have a disadvantage in meta-analytic and other contexts in which comparisons of regression coefficients are made across studies. To the extent the sample means differ across studies, it can be misleading to compare lower order regression coefficients even if identical measures, regression equations, and coding systems are used in each study. Note, however, that the highest order interaction is *not* affected by this problem.

Testing Overall Effects: Multiple *df* Tests

In traditional ANOVA, we have one overall test of each main effect, each two-way interaction, etc.[5] In a design with a categorical variable having $G = 3$ levels, the test of the significance of the categorical variable has $G - 1 = 2$ degrees of freedom. When the overall test of the

4. With modern computer programs, multicollinearity almost never has an effect on the estimate of the standard error of the highest order interaction.

5. When researchers have several strong a priori hypotheses, some authors (e.g., Judd & McClelland, 1989; Rosenthal & Rosnow, 1985) recommend using contrast codes and directly reporting the 1 *df* test of each hypothesis.

$G = 3$ group categorical variable is translated into multiple regression analysis, the test of significance of the categorical variable is actually an omnibus test of whether the b_3C_1 plus the b_4C_2 terms of Equation 2 taken together contribute significant prediction to the outcome, over and above all other predictors in the equation. This test of the joint contribution of the two predictors C_1 and C_2 that represent the categorical variable has 2 degrees of freedom, just as in ANOVA.

To illustrate, consider testing the joint contribution of the two codes carrying the group effect in Equation 2, which is reproduced below:

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_6XC_1 + b_7XC_2 + b_8WC_1 + b_9WC_2 + b_{10}XWC_1 + b_{11}XWC_2 \quad (2)$$

Several reduced models must be estimated that omit the terms of interest. These reduced models, in turn, are each compared with the full regression model given in Equation 2.

Reduced model to test the first-order effect of group,

$$\hat{Y} = b_0 + b_1X + b_2W + b_5XW + b_6XC_1 + b_7XC_2 + b_8WC_1 + b_9WC_2 + b_{10}XWC_1 + b_{11}XWC_2 \quad (4)$$

Reduced model to test the Group $\times$ X interaction,

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_8WC_1 + b_9WC_2 + b_{10}XWC_1 + b_{11}XWC_2 \quad (5)$$

Reduced model to test the Group $\times$ W interaction,

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_6XC_1 + b_7XC_2 + b_{10}XWC_1 + b_{11}XWC_2 \quad (6)$$

Reduced model to test the Group $\times$ X $\times$ W interaction,

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_6XC_1 + b_7XC_2 + b_8WC_1 + b_9WC_2 \quad (7)$$

Following Cohen and Cohen (1983, chap. 4), the R^2 from the full model is compared with the R^2 from the reduced model using Equation 8 (below) to test the gain in prediction,

$$F = \frac{(R^2_{\text{full}} - R^2_{\text{reduced}})/m}{(1 - R^2_{\text{full}})/(n - k - 1)} \tag{8}$$

with $df = m, n - k - 1$. In this equation, R^2_{full} is the squared multiple correlation from the full model, R^2_{reduced} is the squared multiple correlation from the reduced model, n is the number of subjects, k is the number of predictors in the full regression model not including the intercept (here, 11), and m is the number of terms in the set being tested (here, 2). Thus, reduced Equation 4 is compared with full Equation 2 to test the group effect; Equation 5 is compared with Equation 2 to test the Group $\times$ X interaction; Equation 6 is compared with Equation 2 to test the Group $\times$ W interaction; and Equation 7 is compared with Equation 2 to test the Group $\times$ X $\times$ W interaction. These tests can be easily conducted using hierarchical entry procedures in programs such as SPSS REGRESSION or SAS PROC REG. They are also calculated directly by general linear model programs such as SAS PROC GLM. Note that the outcomes of these tests are independent of the choice of coding system for the categorical variable (i.e., all four coding systems lead to the same results).

Example 1: Simulated Data

To provide our first example, the SAS 5.18 random number generator was used to create an artificial data set. Two continuous variables (X and W) representing measured individual difference variables were each constructed to have mean = 0 (centered) and to be normally distributed. X and W were also constructed to be highly correlated ($r = .53$). Three groups were then created ($n_1 = 60$, $n_2 = 220$, $n_3 = 120$). Such discrepant sample sizes are most often found when the groups represent natural categories or experimental treatments that differ widely in cost or difficulty of implementation. Different regression equations were generated in each of the three groups. The difference in the regression coefficients for the XW terms among the three groups produces a three-way interaction in the population.

$$\hat{Y} = 0.7730 + 3.1393X - 2.1514W - 2.1639XW \text{ in Group 1} \tag{9}$$

$$\hat{Y} = 1.1869 - 0.5477X - 0.2865W + 0.9912XW \text{ in Group 2} \tag{10}$$

$$\hat{Y} = -3.3034 - 0.8533X + 2.2922W + 1.0602XW \text{ in Group 3} \tag{11}$$

Normally distributed random error of measurement was added to each predicted value so that the data closely approximated the usual assumptions of the regression model (i.e., residuals are normally and independently distributed with constant variance). Our focus with Example 1 will be to explore the effects of the four methods of coding the group variable: dummy codes, weighted effects codes, unweighted effects codes, and contrast codes.

Interpretation of regression coefficients

Dummy-coded analysis. The first analysis of the data used the dummy codes presented in Table 1(A), Group 3 as Base. The regression equation was structured as in Equation 2. This model was estimated using the simulated data set and the results are presented in the first two columns in Table 2. These columns give the unstandardized regression coefficients, standard errors, t tests, and significance levels of Equation 2.

As we showed earlier, the interpretation of the regression coefficients can be derived by substituting the values of the dummy codes corresponding to each group into Equation 2. For Group 1,

$$\hat{Y} = b_0 + b_1X + b_2W + b_3(1) + b_4(0) + b_5XW + b_6X(1) + b_7X(0) + b_8W(1) + b_9W(0) + b_{10}XW(1) + b_{11}XW(0)$$

Simplifying and collecting terms reduces the equation to

$$\hat{Y} = (b_0 + b_3) + (b_1 + b_6)X + (b_2 + b_8)W + (b_5 + b_{10})XW \quad (12a)$$

For Group 2, the equation reduces to

$$\hat{Y} = (b_0 + b_4) + (b_1 + b_7)X + (b_2 + b_9)W + (b_5 + b_{11})XW \quad (12b)$$

For Group 3 (with values of 0 on both dummy codes), the equation reduces to

$$\hat{Y} = b_0 + b_1X + b_2W + b_5XW \quad (12c)$$

From Equation 12c we see that for dummy coding, the regression coefficient b_1 in the full regression equation (Equation 2) gives the regression of Y on X in Group 3, the comparison group. The regression of Y on X

Table 2

Results of Regression Analyses for Three Coding Systems: Simulated Data

	Dummy codes		Weighted effects		Unweighted effects		Contrast codes	
	b (SE_b)	t p	b (SE_b)	t p	b (SE_b)	t p	b (SE_b)	t p
b_0	−3.30	−1.97	−0.22	−0.26	−0.45	−0.45	−0.45	−0.45
	(1.67)	.05	(0.85)	.79	(1.00)	.65	(1.00)	.65
b_1X	−0.85	−0.50	−0.09	−0.10	0.58	0.50	0.58	0.50
	(1.69)	.62	(0.91)	.92	(1.16)	.61	(1.16)	.61
b_2W	2.29	2.91	0.21	0.50	−0.05	−0.10	−0.05	−0.10
	(0.78)	.004	(0.41)	.61	(0.50)	.92	(0.50)	.92
Group				$F(2, 388) = 2.58, p = .08$				
b_3C_1	4.08	1.46	1.00	0.62	1.22	0.75	4.28	2.05
	(2.79)	.14	(2.05)	.63	(1.63)	.45	(2.09)	.04
b_4C_2	4.49	2.24	1.41	0.81	1.63	1.38	−0.41	0.17
	(2.00)	.03	(0.78)	.07	(1.18)	.17	(2.48)	.87
b_5XW	1.06	1.54	0.54	1.61	−0.04	−0.10	−0.04	−0.10
	(0.69)	.12	(0.33)	.11	(0.39)	.92	(0.39)	.92
Group × X				$F(2, 388) = 0.82, p = .44$				
b_6XC_1	3.99	2.21	3.23	1.28	2.55	1.28	2.15	0.94
	(3.30)	.23	(2.53)	.20	(2.00)	.20	(2.28)	.35

b_7XC_2	0.31	0.15	−0.46	−0.55	−1.12	−0.84	3.69	1.21
	(2.05)	.88	(0.83)	.58	(1.34)	.40	(3.05)	.23
Group × W				$F(2, 388) = 6.05, p = .003$				
b_8WC_1	−4.44	−3.19	−2.36	−2.26	−2.10	−2.54	−3.51	−3.48
	(1.39)	.002	(1.04)	.03	(0.83)	.01	(1.01)	.0006
b_9WC_2	−2.58	−2.72	−0.49	−1.31	−0.24	−0.41	−1.86	−1.48
	(0.95)	.007	(0.38)	.19	(0.58)	.68	(1.26)	.14
Group × X × W				$F(2, 388) = 6.17, p = .002$				
$b_{10}XWC_1$	−3.22	−2.99	−2.70	−3.51	−2.12	−3.46	−1.64	−1.98
	(1.08)	.003	(0.77)	.0005	(0.61)	.0006	(0.83)	.05
$b_{11}XWC_2$	−0.07	−0.09	0.45	1.48	1.03	2.25	−3.16	−3.39
	(0.81)	.93	(0.31)	.14	(0.46)	.03	(0.93)	.0008

Note. The regression coefficients correspond to those in Equation 2:

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C_1 + b_4C_2 + b_5XW + b_6XC_1 + b_7XC_2 + b_8WC_1 + b_9WC_2 + b_{10}XWC_1 + b_{11}XWC_2$$

The four entries for each coefficient for each coding system are: Row 1, estimate of unstandardized regression coefficient and t value; Row 2, standard error of unstandardized regression coefficient (in parentheses) and p value.

in Group 1 is $(b_1 + b_6)$ for Equation 12a. The regression of Y on X in Group 2 is given as $(b_1 + b_7)$. The three regression equations (12a, 12b, 12c) are simple regression equations showing the regression of the dependent variable on the continuous predictors at specific values of the other predictors (here, code variables that signify group membership). The slopes of these regressions, i.e., b_1, $(b_1 + b_6)$, and $(b_1 + b_7)$ for the regression of Y on X for Groups 3, 1, and 2, respectively, are simple slopes. The simple slopes are completely comparable to the ANOVA simple effects for the effect of personality variable X on the dependent variable within each group ("at each level of group").

The four coefficients present in simple slope Equation 12c have direct interpretations that relate to the values of the intercept, simple slopes of Y on X and W, and the XW interaction in Group 3, the comparison group. Note that the values of the unstandardized regression coefficients reported in Table 2 for the intercept, $b_0 = -3.30$, the slope for X, $b_1 = -0.85$, the slope for W, $b_2 = 2.29$, and the XW interaction, $b_5 = 1.06$, precisely equal the corresponding values in the individual equation for Group 3 (Equation 11) presented above.

Turning now to Group 1 (Equation 12a), the intercept for Group 1 is $(b_0 + b_3)$. b_3, then, is the difference between the intercepts for Group 1 and Group 3, and this value (4.08) is equal to the difference between the Group 1 and Group 3 intercepts from the individual equations above (Equations 9 and 11). $(b_1 + b_6)$ is the simple slope of Y on X in Group 1, $(b_2 + b_8)$ is the simple slope of Y on W in Group 1, and $(b_5 + b_{10})$ represents the magnitude of the XW interaction in Group 1. The b_3, b_6, b_8, and b_{10} coefficients are interpreted respectively as differences between the intercept (4.08), slope for X (3.99), slope for W (-4.44), and XW interaction (-3.22) in Group 1 and the corresponding effect in the comparison group (Group 3). Again, the estimated values in the regression analysis are identical to the corresponding differences between the individual equations for Group 1 (Equation 9) and Group 3 (Equation 11) presented above.

The interpretations for b_4, b_7, b_9, and b_{11} (see Equation 12b) exactly parallel those discussed in the preceding paragraph (i.e., intercept, X slope, W slope, XW interaction, respectively), except that they refer to differences between Group 2 and Group 3 rather than to comparisons between Group 1 and Group 3.

Weighted effects coded analysis. Developing interpretations for regression coefficients generated in a weighted effects coded analysis proceeds in the same manner as with dummy codes. Weighted effect codes for the three-group case were presented in Table 1(C), under "General Form," with the specific values of these codes for the present simulation example in which the group sizes are 60, 220, and 120 also presented in Table 1(C), under "Illustration." We substitute the values of C_1 and C_2 for each group into Equation 2, the full regression model and collect terms. The full algebra is presented in the Appendix for interested readers.

The result of the algebra shows that b_0 is the weighted average of the three individual group intercepts. A numeric example illustrating this result is provided by the analysis of the simulated data with weighted effects codes (see Table 2, columns under "Weighted Effects" heading). The value of $b_0 = -.22$ is, in fact, equal to the weighted average of the intercepts from the individual group equations (Equations 8, 9, 10),

$$\begin{aligned} b_0 &= [.7730(60) + 1.1869(220) - 3.3034(120)]/[60 + 220 + 120] \\ &= -.22 \end{aligned}$$

Both the algebra and the numeric example show that b_3 is the difference between the Group 1 intercept and the weighted average of intercepts. Similarly, b_4 is the difference between the Group 2 intercept and the weighted average of intercepts.

Performing the same series of steps on the X slope, W slope, and XW interaction portions of the full regression equation provides analogous interpretations for the rest of the coefficients in this analysis. b_1 is the weighted average of the X slopes for the individual groups, b_2 is the weighted average of the W slopes for the individual groups, and b_5 is the weighted average of the XW interactions for the individual groups. b_6 is the difference between the X slope for Group 1 and the weighted average of individual slopes, b_8 is the difference between the W slope for Group 1 and the weighted average of the W slopes, and b_{10} is the difference between the Group 1 XW interaction and the weighted average of the XW interactions. b_7 is the difference between the X slope in Group 2 and the weighted average of the X slopes in the individual groups, b_9 is the difference between the W slope for Group 2 and the weighted average of the W slopes, and b_{11} is the difference between the XW interaction for Group 2 and the weighted average of the XW interactions across the three groups.

For completeness, we can also calculate the corresponding values for

Group 3, again based on the algebra in the Appendix. The difference between the Group 3 intercept and the weighted average of the intercepts for the three groups is $(b_3C_1 + b_4C_2) = [b_3(-n_1/n_3) + b_4(-n_2/n_3)$. The value $[b_6(-n_1/n_3) + b_7(-n_2/n_3)]$ is the difference between the X slope in Group 3 and the weighted average of the X slopes in the individual groups. The difference between the W slope and XW interaction in Group 3 and the corresponding weighted averages across the three groups are $[b_8(-n_1/n_3) + b_9(-n_2/n_3)]$ for W and $[b_{10}(-n_1/n_3) + b_{11}(-n_1/n_3)]$, respectively.

Unweighted effects codes. Comparison of unweighted with weighted effects codes (see Table 1) indicates that the only difference is that unweighted effects codes have a value of $[-1, -1]$ for Group 3 rather than $[-n_1/n_3, -n_2/n_3]$, which provides the weighting by group size. To develop the interpretation for unweighted effects codes, we use the same logic that was presented above for weighted effects codes with one exception: $[-1, -1]$ is substituted in for the values of C_1 and C_2 in Group 3. The result is that each n_i (the sample size in Group i) in the equations presented for weighted effects codes is replaced by 1, greatly simplifying the algebra (see Appendix). The result is that, for unweighted effects codes, b_0 is the unweighted mean of the intercepts of the regressions in the individual groups,

$$b_0 = (I_1 + I_2 + I_3)/3$$

For example, b_0 is -0.45 in the regression analysis for the full model (Equation 2) presented in Table 2 in the columns under the "Unweighted Effects" heading. This value equals the unweighted mean of the intercepts of the three individual group regression equations (Equations 9, 10, 11),

$$b_0 = [0.7730 + 1.1869 + (-3.3034)]/3 = -0.45$$

Correspondingly, b_1 is the unweighted mean of the X slopes in the individual groups, b_2 is the unweighted mean of the W slopes in the individual groups, and b_5 is the unweighted mean of the XW interactions in the individual groups. The other regression coefficients in the equation have the same interpretation as they did in unweighted effects coding, except that they represent the difference between the statistic for the specified group (e.g., X slope for Group 1) and the *unweighted* mean of the corresponding statistics for the three groups (e.g., unweighted mean of X slopes for individual groups).

Contrast codes. The sets of contrast codes that are most commonly used, such as those presented in Table 1(D), are not weighted by group size. Consequently, the interpretation of b_0, b_1, b_2, and b_5 exactly parallel their respective interpretations in unweighted effect codes. The contrast code analysis is illustrated using the simulated data set in the far-right section of Table 2 under the "Contrast Codes" headings. $b_0 = -0.45$ is the unweighted mean of the intercepts, $b_1 = 0.58$ is the unweighted mean of the X slopes, $b_2 = -0.05$ is the unweighted mean of the W slopes, and $b_5 = -0.04$ is the unweighted mean of the XW interactions in the three individual groups, just as in the unweighted effects code analysis. b_3 represents the difference between the intercept in Group 3 and the unweighted mean of the intercepts in Groups 1 and 2. This can be seen by using the values from Equations 9, 10, 11:

$$b_3 = -3.3034 - [(0.7730 + 1.1869)/2] = -4.28$$

Similarly, b_6 represents the difference between the X slope of Group 3 and the unweighted mean of the X slopes of Groups 1 and 2. b_8 represents the difference between the W slope of Group 3 and the unweighted mean of the W slopes of Groups 1 and 2. b_{10} represents the difference between the XW interaction term in Group 3 and the unweighted mean of the XW interaction terms in Groups 1 and 2. Finally, b_4 represents the difference between the intercepts, b_7 the difference between the X slopes, b_9 the difference between the W slopes, and b_{11} the difference between the XW interaction terms in Groups 2 and 1.

An interpretational caveat. We remind the reader of the important caveat presented earlier regarding the interpretation of regression coefficients in models containing interactions. This caveat applies for each of the four coding systems. All terms except the highest order interaction(s) (here, XWC_1 and XWC_2) are conditional effects that are interpreted at the value of 0 for the variables not involved in the term (Aiken & West, 1991; Cohen, 1978). For example, b_1 represents the slope for X when $C_1 = 0$, $C_2 = 0$, and $W = 0$. For dummy coding and centered W, this effect is interpreted in Group 3 (comparison group) at the mean value (0) for W. For unweighted effects coding (or contrast coding) and centered W, this effect is interpreted at the unweighted mean of the three groups and centered W.

Significance tests

Two different outcomes for significance testing are apparent for the four coding systems that are included in Table 2. First, identical results are produced for the omnibus 2 degree of freedom tests of Group, Group $\times$ X, Group $\times$ W, and Group $\times$ $X \times W$ for each of the coding systems. The same total variance for each of the effects involving group is accounted for by each of the coding systems. Second, the interpretations of the corresponding regression coefficients differ across coding systems. For example, $b_2 = 2.29$ is the W slope in Group 3 for dummy coding, $b_2 = 0.21$ is the weighted mean of the individual W slopes for the three groups for weighted effects coding, and $b_2 = -0.05$ is the unweighted mean of the individual W slopes for the three groups in unweighted effects coding and contrast coding. Given that the same regression coefficient (e.g., b_2) answers a different question in each coding system, it is not surprising that t tests of these regression coefficients produce different results. For example, the test of b_2 is for dummy coding, $t(388) = 2.91, p < .004$; for weighted effect coding, $t(388) = 0.21, ns$; and for unweighted effect and contrast coding, $t(388) = -0.10, ns$. The lesson is clear: Each coding system represents a different partitioning of the total variance associated with the overall 2 *df* group effect.[6]

Issues in the choice of a coding system

In regression equations that parallel familiar complete factorial ANOVA models, each of the coding systems provides directly interpretable results. The choice of coding system depends on the specific questions of the researcher.

1. If the researcher's primary interest is in comparing the parameters (means, slopes, interactions) of each treatment group with those of a specific comparison group, dummy coding gives these results directly. For example, a researcher investigating the interaction of IQ and

6. Researchers wishing to prepare an ANOVA table can easily do so. Some computer programs (e.g., SAS PROC GLM) provide the F tests for each first-order effect, two-way interaction, and three-way interaction directly. Regression programs typically provide t tests of each of the 1 df terms. These t tests can easily be converted since $F = t^2$ for 1 df tests.

psychiatric diagnosis (schizophrenic, depressive, normal) in predicting cognitive processing of information could benefit by using dummy codes. If the normal group is chosen as the base group, then the regression coefficients associated with dummy codes and their interaction compare each of the psychiatric groups with the normal group.

2. If the researcher is interested in producing estimates that most closely parallel those of ANOVA, the researcher must choose between unweighted effect codes, weighted effect codes, and contrast codes. Recall that main effects and lower order interactions in ANOVA are more properly considered to be average effects. In regression analysis, parallel results are obtained only when the values of the codes C_1, C_2, etc. for a group variable are all equal to 0 at the unweighted or weighted mean of the sample. Then in Equation 2, b_0 represents the grand (or average mean), b_1 represents the average X slope, b_2 represents the average W slope, and b_5 represents the average interaction. To illustrate, consider the Rotter and Mulry (1965) study investigating the effects of locus of control and task description on subject's decision time described at the beginning of this article. If Rotter and Mulry predicted both a first-order (average) effect of locus of control and a Locus of Control × Task Description interaction, effect coding (or contrast coding) would provide directly interpretable tests of both predictions.

The choice between unweighted effects, weighted effects, and contrast codes depends on the nature of the group variable and the questions of interest to the researcher.

When the groups represent experimental treatments, differences in sample sizes among groups do not represent meaningful differences in the proportion of each group in the population. Rather, differences in group size will typically reflect factors such as peculiarities of the randomization process or experimenter decisions to include more or fewer subjects because of considerations of the importance, cost, or difficulty in mounting each treatment condition. Under these circumstances, unweighted effects codes that weight each group equally in the estimation of all effects in the equation will be preferred to weighted effects codes. The interpretation of the results from the unweighted effects analysis will then most closely parallel the interpretation in ANOVA when applied to experiments. For example, statements about the average slope of a continuous variable X in the sample will represent the unweighted average of the slopes in each of the individual treatment groups, again given the caveat that $W = 0$ (mean of W in centered solution).

When the groups represent experimental treatments and the researcher has several a priori hypotheses involving contrasts among the groups, contrast codes are preferred. The interpretation of regression coefficients associated with the continuous variables and their interaction is identical to that of unweighted effect codes. The interpretation of the contrast codes and their interactions provide tests of the a priori hypotheses.

When the group variable represents a natural category (e.g., religion) and simple random sampling is used to draw cases from a population, the number of subjects in each specific group (e.g., Catholic), relative to the total number of subjects in the sample, provides an unbiased estimate of the proportion of the population comprised by that group. Under these circumstances, the researcher likely has an interest in estimating relationships that exist in the actual population. Weighted effects codes provide these estimates. In contrast, unweighted effects codes and contrast codes[7] permit less satisfactory generalization of findings if the group variable represents a set of natural categories. The findings can be generalized to a hypothetical population in which each of the groups exists in equal proportions.

When the group sizes are equal, the results of regression using weighted and unweighted effects coding will be identical. When the group sizes are approximately equal, the choice will in practice make little difference in the results (see Example 2 below).

When there are only two groups, unweighted effects codes and contrast codes yield identical significance tests.

3. Researchers wishing to test the overall effects of the group factor and its interactions, as is customary in ANOVA, should conduct the set of tests described in the earlier section, "Testing Overall Effects: Multiple *df* Tests." In each case, the multiple *df* test of the gain in prediction (Equation 8) is used to compare the full regression model to a reduced model in which the terms associated with the effect of interest have been eliminated (e.g., b_3C_1 and b_4C_2 for the overall effect of group in Equation 2). Each first-order effect and interaction involving the group factor must be tested in this manner.

4. Researchers may wish to explore regression models that do not parallel standard complete factorial ANOVA models. Such models introduce additional complexity into the interpretation of the results. For

7. Weighted contrast codes (see Serlin & Levin, 1985) provide appropriate tests when the researcher has several a priori hypotheses about contrasts between groups representing natural categories.

example, consider the following regression equation, which is a generalization of a two-factor analysis of covariance model with one covariate W to be controlled:

$$\hat{Y} = b_0 + b_1X + b_2C_1 + b_3C_2 + b_4XC_1 + b_5XC_2 + b_6W \quad (13)$$

The terms associated with b_0 through b_5 represent a model for an experimental personality design in which the treatment group factor has three levels and X (e.g., need for cognition) is a continuous individual difference variable. W represents another continuous individual difference variable (e.g., IQ) that is to be statistically controlled in the model. In this model, the interpretation of the coefficients for the intercept, X, C_1, C_2, XC_1, and XC_2 terms are exactly as described above for each of the coding systems.[8] However, the effect of W now represents the *pooled within-group regression* coefficient for W. The calculation of the pooled within-group regression coefficient weights the W slope for each individual group by the sum of squares predicted for that group and then computes the weighted mean (see Marascuilo & Levin, 1983, pp. 41, 47–51; Pedhazur, 1982, pp. 438–445). The pooled within-group regression for W will not in general be equal to either the weighted or unweighted average of the individual slopes.

Thus, the interpretation of the b_6 coefficient is distinct from the coefficients for the experimental personality design reflected in b_0 through b_5. Since X interacts with C_1 and C_2, all our prescriptions for interpreting equations containing interactions hold. With dummy codes, b_0 through b_5 are interpreted with respect to the control group. With unweighted effects codes, b_0 through b_5 are interpreted with respect to the unweighted mean of the individual group estimates. With weighted effects codes, b_0 through b_5 are interpreted with respect to the group sample size weighted mean of the individual estimates. With contrast codes, b_0 is interpreted with respect to the unweighted mean of the intercepts, and b_1 through b_5 are interpreted with respect to the specified group contrasts. The covariate W does not interact with C in Equation 13. Hence, the b_6 coefficient is independent of the choice of coding system (i.e., will be constant across coding systems). The interpretation of

8. Because terms have been dropped from the equation, not every b_i in Equation 2 will correspond to the same term as b_i in Equation 12. For example, the b_2 term refers to W in Equation 2 and the first code variable (C) in Equation 12. The regression coefficients that do correspond are those that refer to the same term (e.g., W) in the different equations.

the b_6 coefficient as the pooled within-class regression coefficient is entirely consistent with the familiar regression coefficient for the covariate reported in standard analysis of covariance (ANCOVA) programs.

In general, the interpretation of regression equations like Equation 13 involves distinguishing between the continuous variables (and interactions of continuous variables) that do and do not interact with the set of code variables representing the categorical variable. If a continuous predictor is involved in an interaction with a categorical variable (e.g., X interacts with C_1 and C_2 in Equation 13), then our earlier prescriptions for interpretation as a function of the group coding system hold. If a continuous independent variable does not interact with any other continuous or categorical independent variables in the equation, then the interpretation is *unconditional*, i.e., not dependent on the value of other predictors. The coefficients of continuous independent variables that do not interact with any other variables in the equation are interpreted as pooled within-group regression coefficients, just as in familiar ANCOVA. Interactions of continuous independent variables that do not interact with the categorical variable are interpreted as pooled within-group interactions. Aiken and West (1991) present a full discussion of issues in interpreting interactions between continuous variables. Regression equations that parallel the form of a complete factorial design in ANOVA do not introduce these complexities of interpretation.

Example 2: Self-Esteem, Feedback, and Liking

Given our extensive presentation of the interpretation of coding systems above, we now consider the analysis of an actual data set. We will also introduce methods of graphically presenting the results and post hoc probing of significant interactions.

Overview of study

The example presents a reanalysis of data from an experimental personality design kindly made available to us by Michael Kernis and originally reported in Kernis et al. (1993). This data set concerns the relationships between level of self-esteem, the stability of self-esteem, type of feedback, and various cognitive and emotional response variables. Level of self-esteem, or positivity of one's self-view, was measured via Rosenberg's Self-Esteem Scale (Rosenberg, 1965). To represent stability of self-esteem, Kernis et al. calculated the standard deviation of each individual subject's scores based on eight administrations of a modified version of Rosenberg's Self-Esteem Scale completed at 12-hour intervals over the course of 4 days. We will refer to this variable below as variability of self-esteem to avoid confusion in the interpretation of the direction of the results (i.e., high values represent high variability).

Subjects who had previously been assessed in terms of level and variability of self-esteem were then randomly assigned to either a positive or negative feedback condition. In both feedback conditions, subjects recited a passage from Kurt Vonnegut's novel *Cat's Cradle* in front of a one-way mirror and then received an evaluation of their social skills from a fictitious evaluator supposedly positioned behind the mirror. In the positive feedback condition, the evaluation stated that the observer felt that the subject seemed very socially skillful. In the negative feedback condition, the evaluation stated that the observer felt that the subject did not seem very socially skillful. Subjects then completed a measure of their own emotional reactions and made a number of ratings of the evaluator. Here, we report the reanalysis of one of these ratings, liking for the evaluator, for the 97 subjects who had complete data ($n_{pos} = 50$ for the positive feedback condition; $n_{neg} = 47$ for the negative feedback condition).

Structuring the regression model and specifying a coding system

The independent variables in each analysis were level of self-esteem (continuous), variability of self-esteem (continuous), and feedback condition (a two-level categorical variable). Each continuous predictor variable was centered prior to analysis. For the purposes of illustration, separate analyses were conducted using each of three codings of the categorical variable: (*a*) dummy coding, (*b*) unweighted effects coding, and (*c*) weighted effects coding. Equation 1, which contains all first-order effects, two-way interactions, and three-way interactions among the level (X), variability (W), and feedback (C) variables, was estimated. Equation 1 is reproduced below:

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C + b_4XW + b_5XC + b_6WC + b_7XWC \quad (1)$$

Given that the group variable was an experimental manipulation and the regression model paralleled a complete factorial ANOVA design, unweighted effects codes (equivalent to contrast codes multiplied by a constant in the two-group case) are preferred since this system provides a coherent set of estimates in which the two treatment groups are weighted equally. Dummy codes provide useful information about differences between the two groups when the two measured variables have a value of 0 (i.e., at the means of the measured variables); this interpretation will be particularly useful in post hoc tests. Weighted effects codes are less useful in this experimental context, but are presented here for completeness. In addition, given that the group sizes are approximately equal, the results of weighted effect coding are expected to be approximately equal to those of unweighted effects coding. Results of regression analyses with these three coding systems are given in Table 3.

Interpreting the unstandardized regression coefficients

The interpretation of the unstandardized regression coefficients follows the same framework as in Example 1. However, note that the two-group case permits one major simplification in interpretation relative to the three or more group case. Recall that the multiple *df*, omnibus tests of the first-order effect, and interactions involving the group factor are identical across coding systems. In the two-group case, the 1 *df* test of each of the effects involving the group factor is equivalent to the omnibus test of the gain in prediction. Thus, in the two-group case, *the* t *tests and associated significance levels* for each of the effects involving group will be identical across coding systems. This can be seen in Table 3 in the t tests and p values for each of the effects involving group: Feedback (C), $t(89) = 11.88$, $p = .0001$, Feedback × Level of Self-Esteem, $t(89) = 2.79$, $p = .006$, Feedback × Variability of Self-Esteem, $t(89) = 0.67$, *ns*, and Feedback × Level × Variability of Self-Esteem, $t(89) = 3.09$, $p = .003$. In contrast, the t tests and p values for the corresponding terms involving C_1 and C_2 in Example 1 (see Table 2) differ across coding systems. Thus, comparison of the analogous effects in the three-group (Example 1) and two-group (Example 2) data sets verifies that this simplification only occurs in the two-group case.

At the same time, note in Table 3 that the value of the corresponding *unstandardized regression coefficients* for the effects involving group differ across the three coding systems. For example, consider the value of the feedback first-order effect, b_3: $b_3 = 6.910$ for dummy coding,

Table 3
Results of Regression Analyses for Three Coding Systems: Reanalysis of Kernis et al. (1993) Data

	Dummy codes		Weighted effects		Unweighted effects	
	b (SE_b)	t p	b (SE_b)	t p	b (SE_b)	t p
b_0 intercept	7.839 (0.425)	18.45 .0001	11.401 (0.290)	39.27 .0001	11.294 (0.291)	38.84 .0001
b_1 SE level (X)	−0.130 (0.052)	−2.51 .01	−0.022 (0.037)	−0.60 .55	−0.026 (0.037)	−0.50 .50
b_2 SE variable (W)	−0.005 (0.101)	−0.05 .96	0.049 (0.078)	0.62 .53	0.047 (0.078)	0.61 .55
b_3 feedback (C)	6.910 (0.582)	11.88 .0001	3.348 (0.282)	11.88 .0001	3.455 (0.291)	11.88 .0001
$b_4X \times W$	−0.035 (0.014)	−2.43 .02	−0.006 (0.009)	−0.63 .53	−0.007 (0.009)	−0.72 .47
$b_5X \times C$	0.209 (0.075)	2.79 .006	0.101 (0.036)	2.79 .006	0.104 (0.037)	2.79 .006
$b_6W \times C$	0.104 (0.155)	0.67 .50	0.050 (0.075)	0.67 .50	0.052 (0.078)	0.67 .50
$b_7X \times W \times C$	0.056 (0.018)	3.09 .003	0.027 (0.009)	3.09 .003	0.028 (0.009)	3.09 .003

Note. The regression coefficients correspond to those in Equation 12:

$$\hat{Y} = b_0 + b_1X + b_2W + b_3C + b_4XW + b_5XC + b_6WC + b_7XWC$$

The four entries for each coefficient for each coding system are: Row 1, estimate of unstandardized regression coefficient and t value; Row 2, standard error of unstandardized regression coefficient (in parentheses) and p value.

$b_3 = 3.348$ for weighted effects coding, and $b_3 = 3.455$ for unweighted effects coding. This value represents the difference (6.910) between predicted means in the negative and positive feedback groups for dummy coding, the difference (3.348) between the predicted mean of the negative feedback group and the predicted unweighted mean of the two treatment groups for unweighted effects coding, and the difference (3.455) between the predicted mean of the negative feedback group and the predicted weighted mean of the two treatment groups for weighted effects coding. In all three coding systems, the group effects are interpreted at the point where the values of level (X) and variability (W) of self-esteem are both 0. When the variables X and W are centered, this point is the mean value of the two continuous independent variables.

In contrast, both the regression coefficients and the t tests for all effects involving only the continuous variables differ across the coding systems. To illustrate, consider b_1, the regression coefficient for the first-order effect of level of self-esteem. In Table 3, in the dummy-coded analysis, $b_1 = -0.130$, $t(89) = -2.51, p = .014$, represents the slope of level of self-esteem in the comparison group (the negative feedback condition). In the weighted effects coding analysis, $b_1 = -0.022$, $t(89) = -0.60$, *ns*, represents the weighted mean of the individual slopes for level of self-esteem in the two feedback conditions. Finally, in the unweighted effects coding analysis, $b_1 = -0.026$, $t(89) = -0.50$, *ns*, represents the unweighted mean of the individual slopes for level of self-esteem in the two feedback conditions. Regardless of the coding system for feedback, the regression coefficient for level of self-esteem is interpreted at 0 in the coding system for the categorical variable.

Table 3 shows that for the unweighted effects codes, which most closely parallel the typical ANOVA results, we have a significant first-order effect for feedback, modified by a Feedback × Level of Self-Esteem two-way interaction, and a Feedback × Level × Variability of Self-Esteem three-way interaction. Again, very similar results are found for weighted effects coding, as expected given the approximately equal sample sizes of the two groups ($n_1 = 50$; $n_2 = 47$). To further understand these significant interactions, it is useful to present the results graphically and to conduct tests of simple effects.

Graphical display of results

There are several methods of graphically displaying the results of complex interactions between categorical and continuous variables (see Cleveland, 1993, 1994). We present two particularly useful ones below: traditional two-dimensional graphs (co-plots) and newer three-dimensional graphs (perspective plots).

Two-dimensional graphs. To plot the three-way interaction between a categorical and two continuous variables, separate graphs representing the interactions between the two continuous variables are constructed for each treatment group. One of the continuous variables, here level of self-esteem, is placed on the X-axis and the predicted value of $Y(\hat{Y})$, here predicted liking for the evaluator, is placed on the Y-axis. For the other continuous variable W, here variability of self-esteem, the researcher must choose specific values at which a regression line will be plotted. When meaningful values of W exist a priori (e.g., on a measure of depression, a score typical of a normal population, a clinical cutoff score, and a score typical of a clinical population based on normative data), these values should be chosen. In the absence of a priori meaningful values, Cohen and Cohen (1983) suggested the convention of plotting values at one standard deviation below the mean of $W(W_{Low})$, the mean of $W(W_{\text{Mod}})$, and one standard deviation above the mean of $W(W_{\text{High}})$. We will follow Cohen and Cohen's convention in our example.

The next step is to construct the regression equation for each group. Our goal is to construct simple regression equations for the regression of Y on X within one of the groups, each at a specific value of W. Just as the highest order interaction is invariant across the scaling of the continuous predictors, the slopes of the simple regression equations within each group are constant across coding systems.

Constructing simple regression equations is most easily accomplished using the dummy-coded solution. While the identical final results are obtained from the other coding systems, the algebra is more difficult. If we collect terms from the dummy-coded solution involving each continuous variable, we find

$$\begin{aligned}\hat{Y} &= (b_0 + b_3C) + (b_1X + b_5XC) + (b_2W + b_6WC)\\ &\quad + (b_4XW + b_7XWC)\end{aligned} \tag{14}$$

For the negative feedback (comparison) group, we first substitute in the value $C = 0$ and then substitute in the values of all the regression coefficients from the dummy-coded regression analysis in Table 3. This yields the specific regression equation for the negative feedback group in the present data set.

$$\begin{aligned}\hat{Y} &= (7.839 + 0) + (-0.130X + 0) + (-0.005W + 0)\\ &\quad + (-0.035WX + 0)\\ \hat{Y} &= 7.839 + (-0.130)X + (-0.005)W + (-0.035)XW\end{aligned} \tag{15}$$

This equation describes the simple first-order and simple interactive effects of X and W in the negative feedback group.

In a similar manner, we first substitute in the value of $C = 1$ and then once again substitute in the specific value for each regression coefficient from the dummy-coded regression analysis in Table 3 into Equation 14. This yields the specific regression equation for the positive feedback group in the present data set.

$$\hat{Y} = (7.839 + 6.910) + (-0.130 + 0.209)X + (-0.005 + 0.104)W + (-0.035 + 0.056)XW$$

$$\hat{Y} = 14.749 + 0.079X + 0.099W + 0.021XW \tag{16}$$

This equation describes the simple first-order and simple interactive effects of X and W in the positive feedback group.

For centered W, the mean = 0 and the standard deviation of W is 4.039. Following Cohen and Cohen's convention, we plot the simple regression lines at $W_{Low} = -4.039$, $W_{\text{Mod}} = 0$, and $W_{\text{High}} = +4.039$. Substituting these values into the simple regression equation (Equation 15) at three different values of W for the negative feedback group, we find:

For W_{Low}, $\hat{Y} = 7.839 + (-0.130)X + (-0.005)(-4.039) + (-0.035)(-4.039)X,$

$\hat{Y} = 7.859 + 0.011X$ for W_{Low};
$\hat{Y} = 7.839 - 0.130X$ for W_{Mod};
$\hat{Y} = 7.819 - 0.271X$ for W_{High}.

These simple regression lines for the negative feedback group are plotted in Panel A of Figure 1. The lines display the XW interaction in the negative feedback group (i.e., a simple interaction in ANOVA terminology).

We now follow the same procedures, substituting $C = 1$ and the values of the regression coefficients from Table 3 into the simple regression equation (Equation 16) for the positive feedback group. This results in the following simple regression equations at three different values of W for the positive feedback group:

For W_{Low}, $\hat{Y} = 14.749 + 0.079X + 0.099(-4.039) + 0.021(-4.039)X$;

$\hat{Y} = 14.349 - 0.008X$ for W_{Low};
$\hat{Y} = 14.749 + 0.079X$ for W_{Mod};
$\hat{Y} = 15.149 + 0.165X$ for W_{High}.

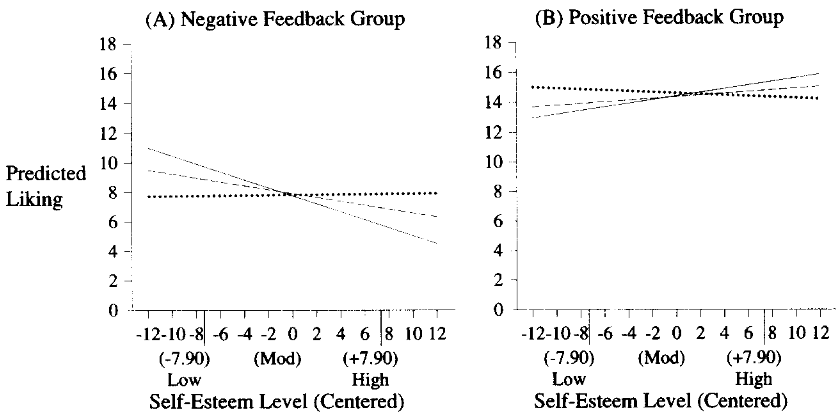

Figure 1
Simple Regression Lines Depicting the Relationship between Level of Self-Esteem and Liking for the Evaluator at Specified Values of Variability of Self-Esteem within Each Feedback Group

Note. Values on the Y-axis represent predicted liking for the evaluator. The solid, dashed, and dotted lines represent high, moderate, and low values of variability of self-esteem, respectively.

These simple regression lines for the positive feedback group are plotted in Panel B of Figure 1. They display the XW interaction in the positive feedback group (i.e., a simple interaction in ANOVA terminology).

Figure 1 shows that in the negative feedback group, the regression of liking on self-esteem becomes progressively more negative as variability of self-esteem increases. In contrast, in the positive feedback group, the regression of liking on self-esteem becomes progressively more positive as variability of self-esteem increases.

Figure 1 presents only one of the possible two-dimensional displays of the data at levels or values of the third variable. Each of the regression equations involving interactions discussed in this article generates a separate regression surface defined by the two continuous predictors X and W for each treatment group. Several different two-dimensional displays of the three-dimensional surface can be constructed. For example, we could construct three panels corresponding to (*a*) W_{Low}, (*b*) W_{Mod}, and (*c*) W_{High} that emphasize the direct comparison of the simple regression lines of Y on X for negative and positive feedback at three different values of variability of self-esteem. In this depiction, the left panel would represent W_{Low} and would display the two simple regression lines:

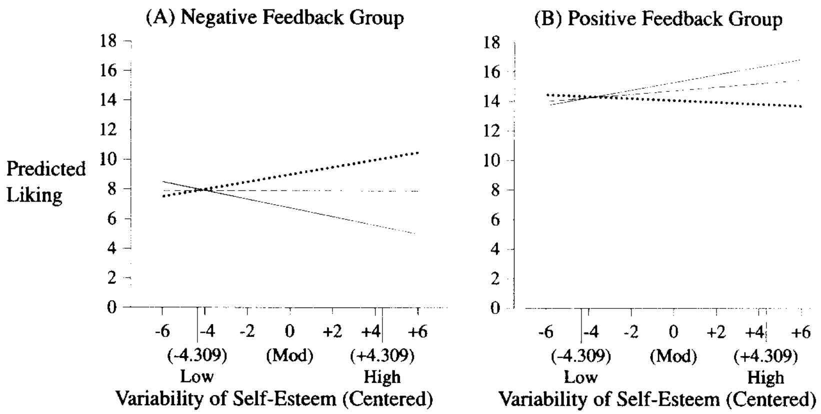

Figure 2
Simple Regression Lines Depicting the Relationship between Stability of Self-Esteem and Liking for the Evaluator at Specified Values of Level of Self-Esteem within Each Feedback Group

Note. Values on the Y-axis represent predicted liking for the evaluator. The solid, dashed, and dotted lines represent high, moderate, and low values of level of self-esteem, respectively.

$$\hat{Y} = 7.859 + 0.011X \text{ for negative feedback;}$$
$$\hat{Y} = 14.349 - 0.006X \text{ for positive feedback.}$$

Alternatively, we could explore the regression of Y on W at three different values of X within each treatment. For level of self-esteem, the mean $= 0$ (centered) and the standard deviation is 7.905. Paralleling the above procedure, these values can be substituted back into Equation 14 and the simple regression lines for Y on W at $X_{\text{Low}} = -7.905$, $X_{\text{Mod}} = 0$, and $X_{\text{High}} = +7.905$ are calculated. These simple regression lines are depicted in Figure 2, Panel A for negative feedback and Panel B for positive feedback.

Each of the different two-dimensional displays can potentially help reveal different aspects of the results. Following the recommendations of Cleveland (1994), the panels should be plotted side by side sharing the same values on the Y axis to facilitate comparisons across groups. This was done in Figures 1 and 2, clearly showing the large first-order effect for feedback group in addition to the different XW interaction within each treatment group. The two-dimensional displays can also be useful for table lookup for the specific values that are plotted. That is, in Figure 1 the predicted value of Y can be read directly from the graph ("looked up") for any value of X for the values of W that are specified.

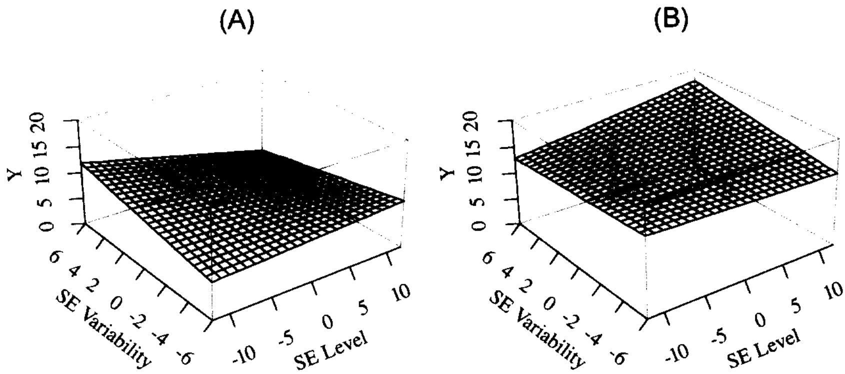

Figure 3
Three-Dimensional Perspective Plot Depicting Regression Surface within Each Feedback Group

Note. SE Level corresponds to level of self-esteem. SE Variability corresponds to variability of self-esteem. Values on the *Y*-axis represent predicted liking for the evaluator. (A) = Negative Feedback Group; (B) = Positive Feedback Group.

Three-dimensional graphs. Another alternative is to directly represent the three-dimensional regression surface separately for each treatment group. Modern graphics programs permit the plotting of regression surfaces using three-dimensional perspective (or wireframe) plots. In Figure 3, the regression surface corresponding to the negative feedback condition is plotted in Panel A and the surface corresponding to the positive feedback condition is plotted in Panel B. These plots can be made with the modern statistical graphics procedures in Mathematica (Wolfram, 1991), S-PLUS (StatSci, 1995), and the S-PLUS Trellis Displays library (MathSoft, 1995), as well as with presentation graphics packages such as Axum (TriMetrix, 1993) and Stanford Graphics (Stanford Graphics, 1994), among others (see Marsh, 1994; Nash, 1994, for reviews of several graphics programs). Following the recommendations of Cleveland (1994), with most data sets the panels should be plotted side by side sharing the same values on the *Y* axis to facilitate comparisons across groups. Unfortunately, not all current graphics packages permit this option. The plots in Figure 3 were done in S-PLUS using the two within-group regression equations (Equations 15 and 16). Other perspectives (viewpoints) on the two regression surfaces can also be plotted.

The three-dimensional perspective plots are particularly useful in visualizing the shape of the regression surface (Cleveland, 1993). Both panels of Figure 3 show that the predicted liking for the evaluator is little affected by feedback for subjects with low values of self-esteem variability, regardless of their level of self-esteem. However, as the variability of self-esteem increases, the relation between the level of self-esteem and liking for the evaluator becomes increasingly negative in the negative feedback condition (Panel A) and increasingly positive in the positive feedback condition (Panel B).

Comment. Both methods of graphical display are useful. The three-dimensional perspective plot is useful in visualizing the shape of the regression surface and permits an overall comparison of the regression surfaces in the two treatment groups. The three-dimensional surface is not useful in table lookup because the viewer cannot accurately track points from the surface to the Y-axis. The two-dimensional graphs present simple regression lines that represent specific values on the regression surfaces. These simple regression lines highlight the differences between specified values on the surfaces and are useful for table lookup. The slopes of these lines are also a focus of post hoc testing procedures.

Post hoc testing procedures

Following a significant three-way interaction, three types of effects may be tested to understand further the nature of the interaction. These are (a) tests of simple slopes of Y on X within groups, (b) tests of simple slopes of Y on W within groups, and (c) tests of group differences at specified pairs of values of the two continuous variables. These tests parallel the practice in ANOVA of post hoc testing of "simple, simple main effects" (i.e., testing group differences on the one factor in an experiment at specified levels on the other factors; see Winer et al., 1991). Each of these three methods of testing is discussed below.

Tests of simple slopes of Y *on* X *within groups.* We derived above the simple regression equations for the regression of Y on X within each of the treatment groups at specified values of W. The regression lines corresponding to these simple regression equations are illustrated in Figure 1. The slopes of each of these equations may be tested against 0 using a general matrix-based procedure developed by Aiken and West (1991, pp. 24–26). More simply, we can use a relatively simple, computer-based method that takes advantage of our understanding of the meaning of regression coefficients when other variables in the equation have a value of 0.

Consider the interpretation of the b_1 coefficient in dummy coding. This coefficient represents the slope of the simple regression of Y on X when $C = 0$ (comparison group, here negative feedback) and $W = 0$ (mean of W). Thus, the b_1 coefficient estimates the simple slope of Y on X for the negative feedback group for W_{Mod} (the mean level of W) and the associated t test assesses whether this value ($b_1 = -.129$) is significantly different from 0. Suppose now we reran the same regression model using the reversed dummy coding system: $C = 0$ for positive feedback and $C = 1$ for negative feedback. Now, the b_1 coefficient estimates the simple slope of Y on X for the positive feedback group for W_{Mod} and the associated t test assesses whether this value is significantly different from 0.

How can we test the simple slopes of Y on X at W_{Low} and W_{High}? Recall that we originally rescaled X and W through centering to simplify the interpretation of the unstandardized regression coefficients. We can now rescale W again so that its value is 0 at W_{Low} or W_{High}. To be clear in this section, we will use W_C rather than W to represent the centered scaling of W that we have discussed up to this point. To make $W_{\text{Low}} = 0$, we create a new variable W_L in which the standard deviation of W (4.039) is added to each subject's score on W_C, that is, $W_L = W_C + 4.039$. This results in scores for W_L that are 0 when $W_C = -4.039$. To make $W_{\text{High}} = 0$, we create a new variable W_H in which the standard deviation of W is subtracted from each subject's score on W_C, that is $W_H = W_C - 4.039$. This results in scores for W_H that are 0 when $W_C = +4.039$. Aiken and West (1991, pp. 18–19) present a further explanation of this procedure.

We now conduct a series of four regression analyses, each using the same regression model (Equation 1) and the centered scores for X. The dummy coding for C is systematically varied (original: negative feedback $= 0$, positive feedback $= 1$; reversed: negative feedback $= 1$, positive feedback $= 0$) and the scores for W_L and W_H replace centered W. Taken together with the tests using W_C (centered) described above, all six simple slopes have been estimated and tested. These results are presented in the top section (A) of Table 4.

Tests of simple slopes of Y *on* W *within groups.* Tests of the simple slope of Y on W at specified values of X illustrated in Figure 2 follow the same general logic. A series of regression equations using the same model and centered W are tested. The dummy coding for C (original; reversed) and values for X_L, X_C, and X_H are systematically varied. This procedure results in six regression equations: In each case, b_2, the coefficient for centered W, is the simple slope for Y on W at the specified values of X and treatment group (feedback condition). The t tests associated with each b_2 coefficient assess the simple slope of Y on W against 0 at the specified values of X and C. These results are presented in the middle section (B) of Table 4.

Tests of group differences at specified values of continuous variables. Recall that when C is dummy coded,[9] b_3 in Equation 1 represents the difference between Group 1 and the comparison group at the value $X = 0$ and $W = 0$. Thus, the test of b_3 in Table 3 represents the difference in the predicted means of the positive feedback and negative feedback groups at X_{Mod} and W_{Mod} when X and W are centered. Using our previous strategy, we can create rescaled X (X_L, X_C, and X_H) and W (W_L, W_C, and W_H) variables that allow us to test the differences between the two treatment groups. Nine separate regression equations representing each of the nine combinations of low, moderate (mean), and high values of X and W are estimated. In each equation, b_3 is the estimate of the difference between the negative and positive feedback conditions at the specified values of X and W. The bottom section (C) of Table 4 presents these estimates and the associated t tests of the null hypothesis that the difference between the two groups at the point specified is 0.

Comment. The three sets of post hoc tests described above closely parallel the three sets of simple effects tests that would follow up a significant $2 \times 2 \times 2$ interaction in ANOVA. In that context, the procedure is known as Fisher's (1935) Least Significant Difference (LSD) method, a method that has its advocates and critics. A conservative researcher wishing fuller protection against an inflated Type I error rate because of the large number of post hoc tests may wish to use the Bonferroni procedure to adjust the alpha level required to reject the null

9. In the two-group case, each of the coding systems produces an equivalent test of the difference between the groups. However, when the treatment is comprised of three or more groups, dummy coding must be used to provide the proper test. For example, in the three-group case, the t test of C_1 tests the difference between Group 1 and the comparison group; the t test of C_2 tests the difference between Group 2 and the comparison group. The difference between Groups 1 and 2 is most easily tested by recoding Group 1 as the comparison group (see Table 1[A], Dummy Codes, Group 1 as Base). The t test of C under this coding system provides the test of the difference between Groups 1 and 2. Each of these t tests assesses the null hypothesis of no difference between two groups at the specified values of X and W.

Table 4
Post Hoc Tests

(A) Tests of simple slopes of *Y* on *X* at values of *W*

		Variability of self-esteem		
		W_{low}	W_{mod}	W_{high}
Feedback				
Negative	b_1	0.009	−0.129	−0.269
	t value	0.11	−2.51	−3.85
	p value	*ns*	.014	.0002
Positive	b_1	−0.008	0.079	0.165
	t value	−0.12	1.46	2.22
	p value	*ns*	*ns*	.029

(B) Tests of simple slopes of *Y* on *W* at values of *X*

		Level of self-esteem		
		X_{low}	X_{mod}	X_{high}
Feedback				
Negative	b_2	0.268	−0.005	−0.277
	t value	1.52	−0.05	−2.29
	p value	*ns*	*ns*	.025
Positive	b_2	−0.070	0.099	0.268
	t value	−0.57	0.84	1.61
	p value	*ns*	*ns*	*ns*

(C) Tests of differences between means at specified values of *X* and *W*

		Level of self-esteem		
		X_{low}	X_{mod}	X_{high}
Variability of self-esteem				
W_{high}	b_3	3.897	7.330	10.762
	t value	3.87	8.70	8.23
	p value	.0002	.0001	.0001
W_{mod}	b_3	5.261	6.910	8.559
	t value	6.29	11.88	10.42
	p value	.0001	.0001	.0001
W_{low}	b_3	6.625	6.490	6.355
	t value	4.81	7.48	6.26
	p value	.0001	.0001	.0001

Note. *X* refers to level of self-esteem and *W* refers to variability of self-esteem. High, moderate, and low refer to values 1 *SD* above the mean, at the mean, and 1 *SD* below the mean, respectively, on each continuous variable.

hypothesis. For example, there are six tests of simple slopes of Y on X (at values W_{Low}, W_{Mod}, W_{High} in each group). The conservative researcher might adopt alpha $= .05/6 = .0083$ for each test, providing control of the error rate for the overall hypothesis that the simple slope of Y on X is 0. The reader should note, however, that control of the hypothesis-wise or even more conservative study-wise Type I error rate is not without its attendant costs in terms of increased Type II error rates (see Cohen, 1994).

Interpretation of Kernis et al. (1993) results

The initial analysis of the liking for the evaluator measure (see Table 3, Unweighted Effects) showed a significant first-order effect of feedback, a two-way interaction of feedback and level of self-esteem, and a three-way interaction of feedback and level and variability of self-esteem.

Follow-up tests of the simple slopes of Y on X (illustrated in Figure 1) showed that in the negative feedback condition, the relation between the subject's level of self-esteem and liking for the evaluator became increasingly negative as the value of the variability of self-esteem variable increased. In the positive feedback condition, the relation between the subject's level of self-esteem and liking for the evaluator was not significant for low or moderate levels of variability of self-esteem, but was significantly positive at high levels of variability of self-esteem.

Tests of the simple slopes of Y on W (illustrated in Figure 2) showed that in the negative feedback condition, the simple regression of liking for the evaluator on the variability of the subject's self-esteem was not significant for low and moderate (mean) values of level of self-esteem, but that there was a significant negative relationship when level of self-esteem was high. In the positive feedback condition, no significant relationship between variability of self-esteem and liking for the evaluator was observed at any of the levels of self-esteem that were considered.

The post hoc tests of group differences at specified values (see Table 4[C]) showed that the evaluator in the high feedback group was significantly better liked at the full range of values of level and variability of self-esteem that were considered. The graphical presentations in Figures 1, 2, and 3 further illustrate and highlight different aspects of these findings. For example, the overall elevation of the set of lines in Figure 1, Panel B (positive feedback) is greater than that in Figure 1, Panel A (negative feedback).

Space limitations do not permit us to illustrate the powerful methods of detecting model misspecification available in multiple regression. Curvilinear effects and interactions involving continuous variables can be detected and the regression model can be respecified to permit tests of such effects. Signs of problems with the data, such as outliers, or with the regression model, such as heteroscedasticity (nonconstant variance) or nonnormality of residuals, can be detected and a variety of corrective procedures can be used. Both graphical methods and formal statistical tests are available. These methods are discussed in more detail in Cook and Weisberg (1994), Hamilton (1992), and Neter, Wasserman, and Kutner (1989).

CONCLUSION

In this article, we have presented a full set of multiple regression-based techniques for the analysis of categorical × continuous variable interactions in between-subject designs. We have considered both experimental personality designs and designs involving natural categories and continuous variables. We have addressed the structuring of regression equations, choice of coding system, and the importance of centering continuous variables. We have considered in detail the interpretation of regression coefficients in each of the coding systems. Finally, we have considered methods for graphically displaying the results and for post hoc testing of simple slopes following significant interactions. The use of the methods outlined in this article provides all of the information available from the use of ANOVA with cutpoints, but without the attendant loss of power and possibility of spurious first-order effects.

Appendix

This appendix shows the algebraic derivation of the meaning of the regression coefficients for weighted effect codes for the three-group case. The specific values of the weighted effect codes are first substituted into Equation 2. For Group 1,

$$\hat{Y} = (b_0 + b_3) + (b_1 + b_6)X + (b_2 + b_8)W + (b_5 + b_{10})XW$$

For Group 2,

$$\hat{Y} = (b_0 + b_4) + (b_1 + b_7)X + (b_2 + b_9)W + (b_5 + b_{11})XW$$

For Group 3,

$$\begin{aligned}\hat{Y} &= [b_0 + (-n_1/n_3)b_3 + (-n_2/n_3)b_4] \\ &+ [b_1 + (-n_1/n_3)b_6 + (-n_2/n_3)b_7]X \\ &+ [b_2 + (-n_1/n_3)b_8 + (-n_2/n_3)b_9]W \\ &+ [b_5 + (-n_1/n_3)b_{10} + (-n_2/n_3)b_{11}]XW\end{aligned}$$

The intercept portions of the equations for Group 1 and Group 2 are $I_1 = b_0 + b_3$ and $I_2 = b_0 + b_4$. Thus

$$b_3 = I_1 - b_0, \tag{A1}$$

and

$$b_4 = I_2 - b_0 \tag{A2}$$

The intercept portion of the equation for Group 3 is

$$I_3 = b_0 + (-n_1/n_3)b_3 + (-n_2/n_3)b_4$$

Substituting the expressions for b_3 and b_4 as functions of the Group 1 and Group 2 intercepts into the Group 3 equation we get

$$I_3 = b_0 + (I_1 - b_0)(-n_1/n_3) + (I_2 - b_0)(-n_2/n_3)$$

Solving for b_0 gives us

$$b_0 = (n_1 I_1 + n_2 I_2 + n_3 I_3)/(n_1 + n_2 + n_3) \tag{A3}$$

Thus, b_0 is the weighted average of the intercepts of the three groups. Equations A1, A2, and A3 thus define b_3, b_4, and b_0, respectively, in terms of the group intercepts. Performing the same series of algebraic steps on the X slope, W slope, and XW interaction portions of the full regression equation provides analogous solutions for the rest of the coefficients in this analysis.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Atkinson, J. W., & Feather, N. T. (1966). *A theory of achievement motivation.* New York: Wiley.

Campbell, J. D. (1993, August). Clarity of the self-concept. Invited address, American Psychological Association, Toronto, Ontario.

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality*, **59**, 143–178.

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart Press (AT&T Bell Laboratories).

Cleveland, W. S. (1994). *The elements of graphing data* (rev. ed.). Summit, NJ: Hobart Press (AT&T Bell Laboratories).

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, **70**, 426–443.

Cohen, J. (1978). Partialed products *are* interactions; partialed vectors *are* curve components. *Psychological Bulletin*, **85**, 858–866.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, **7**, 249–253.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, **49**, 997–1003.

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

Cook, R. D., & Weisberg, S. (1994). *An introduction to regression graphics.* New York: Wiley.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Fisher, R. A. (1935). *The design of experiments.* London: Oliver & Boyd.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, **26**, 797–833.

Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics.* Belmont, CA: Duxbury.

Hardy, M. A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego: Harcourt Brace Jovanovich.

Kernis, M. H., Cornell, C. P., Sun, C-R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal of Personality and Social Psychology*, **65**, 1190–1204.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Krahé, B. (1992). *Personality and social psychology: Towards a synthesis.* London: Sage.

Magnusson, D., & Endler, N. S. (Eds.). (1977). *Personality at the crossroads: Current issues in interactional psychology.* Hillsdale, NJ: Erlbaum.

Marascuilo, L. A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide.* Monterey, CA: Brooks/Cole.

Marquardt, D. W. (1980). You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association*, **75**, 87–91.

Marsh, P. L. (1994). A review and comparison of five graphics programs for PC-DOS and MS-DOS computers. *American Statistician*, **48**, 44–51.

MathSoft, Inc. (1995). *S-PLUS Trellis Graphics user's manual, Version 3.3*. Seattle: Author.

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

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, **114**, 376–390.

Nash, J. C. (1994). Tools for including statistical graphics in applications programs. *American Statistician*, **48**, 52–57.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Reinhart & Winston.

Peixoto, J. L. (1987). Hierarchical variable selection in polynomial regression models. *American Statistician*, **41**, 311–313.

Pitts, S. C., & West, S. G. (1995). *Alternative sampling designs to detect interactions in multiple regression*. Manuscript submitted for publication.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge: Cambridge University Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, **80** (Whole No. 609).

Rotter, J. B., & Mulry, R. C. (1965). Internal versus external control of reinforcements and decision time. *Journal of Personality and Social Psychology*, **2**, 598–604.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, **16**, 209–222.

Serlin, R. C., & Levin, J. R. (1985). Teaching how to derive directly interpretable coding schemes for multiple regression analysis. *Journal of Educational Statistics*, **10**, 223–238.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., Vol. 2, pp. 883–948). New York: Random House.

Stanford Graphics. (1994). *Stanford graphics presentation and analysis user's guide*. Houston: Visual Numerics.

StatSci. (1995). *S-PLUS user's manual, Version 3.3 for Windows*. Seattle: Author.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology*, **79**, 354–459.

Suits, D. B. (1984). Dummy variables: Mechanics v. interpretation. *Review of Economics and Statistics*, **66**, 177–180.

TriMetrix, Inc. (1993). *Axum (Version 2.0): User's guide*. Seattle: Author.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Wolfram, S. (1991). *Mathematica: A system for doing mathematics by computer* (2nd ed.). Reading, MA: Addison-Wesley.